(12) INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(19) World Intellectual Property Organization
International Bureau

(43) International Publication Date
16 August 2001 (16.08.2001)

PCT

(10) International Publication Number
WO 01/59116 A2

(51) International Patent Classification[7]: C12N 15/12, C07K 14/705, 16/28, C12Q 1/68, G01N 33/53, 33/68, A61K 39/395

(21) International Application Number: PCT/US01/04446

(22) International Filing Date: 9 February 2001 (09.02.2001)

(25) Filing Language: English

(26) Publication Language: English

(30) Priority Data:
60/181,301 9 February 2000 (09.02.2000) US

(71) Applicant *(for all designated States except US)*: MILLENNIUM PHARMACEUTICALS, INC. [US/US]; 75 Sidney Street, Cambridge, MA 02139 (US).

(72) Inventors; and

(75) Inventors/Applicants *(for US only)*: GLUCKSMANN, Maria, Alexandra [AR/US]; 33 Summit Road, Lexington, MA 02173 (US). WHITE, David [US/US]; 35 Hollingsworth Avenue, Braintree, MA 02184 (US).

(74) Agents: COULTER, Kathryn, L. et al.; Alston & Bird LLP, P.O. Drawer 34009, Charlotte, NC 28234-4009 (US).

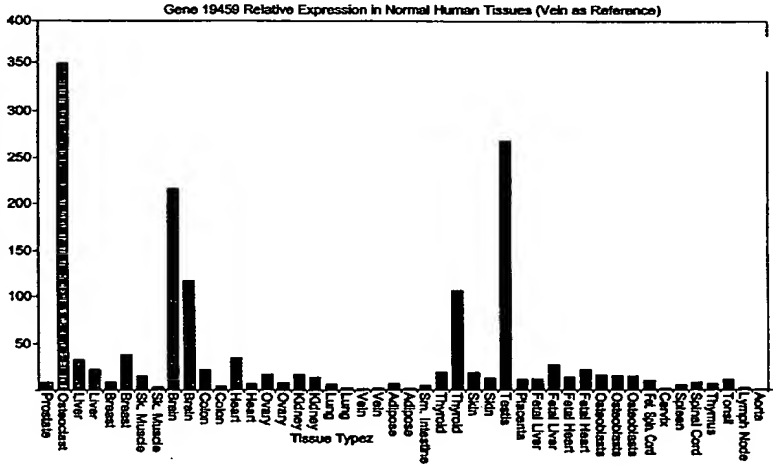
(81) Designated States *(national)*: AE, AG, AL, AM, AT, AT (utility model), AU, AZ, BA, BB, BG, BR, BY, BZ, CA, CH, CN, CR, CU, CZ, CZ (utility model), DE, DE (utility model), DK, DK (utility model), DM, DZ, EE, EE (utility model), ES, FI, FI (utility model), GB, GD, GE, GH, GM, HR, HU, ID, IL, IN, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MA, MD, MG, MK, MN, MW, MX, MZ, NO, NZ, PL, PT, RO, RU, SD, SE, SG, SI, SK, SK (utility model), SL, TJ, TM, TR, TT, TZ, UA, UG, US, UZ, VN, YU, ZA, ZW.

(84) Designated States *(regional)*: ARIPO patent (GH, GM, KE, LS, MW, MZ, SD, SL, SZ, TZ, UG, ZW), Eurasian patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European patent (AT, BE, CH, CY, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE, TR), OAPI patent (BF, BJ, CF, CG, CI, CM, GA, GN, GW, ML, MR, NE, SN, TD, TG).

Published:
— *without international search report and to be republished upon receipt of that report*

[Continued on next page]

(54) Title: 19459 PROTEIN, A NOVEL SEVEN TRANSMEMBRANE PROTEIN



(57) Abstract: The present invention relates to a newly identified seven-transmembrane protein, potentially a receptor belonging to the superfamily of G-protein-coupled receptors. The invention also relates to polynucleotides encoding the protein. The invention further relates to methods using the polypeptides and polynucleotides as a target for diagnosis and treatment in 19459 protein-mediated or -related disorders. The invention further relates to drug-screening methods using the polypeptides and polynucleotides to identify agonists and antagonists for diagnosis and treatment. The invention further encompasses agonists and antagonists based on the polypeptides and polynucleotides. The invention further relates to procedures for producing the polypeptides and polynucleotides.

WO 01/59116 A2

For two-letter codes and other abbreviations, refer to the "Guidance Notes on Codes and Abbreviations" appearing at the beginning of each regular issue of the PCT Gazette.

19459 PROTEIN, A NOVEL SEVEN TRANSMEMBRANE PROTEIN

FIELD OF THE INVENTION

The present invention relates to a newly identified seven-transmembrane protein, potentially a receptor belonging to the superfamily of G-protein-coupled receptors. The invention also relates to polynucleotides encoding the protein. The invention further relates to methods using the polypeptides and polynucleotides as a target for diagnosis and treatment in 19459 protein-mediated or -related disorders. The invention further relates to drug-screening methods using the polypeptides and polynucleotides to identify agonists and antagonists for diagnosis and treatment. The invention further encompasses agonists and antagonists based on the polypeptides and polynucleotides. The invention further relates to procedures for producing the polypeptides and polynucleotides.

BACKGROUND OF THE INVENTION

G-protein coupled receptors

G-protein coupled receptors (GPCRs) constitute a major class of proteins responsible for transducing a signal within a cell. GPCRs have three structural domains: an amino terminal extracellular domain, a transmembrane domain containing seven transmembrane segments, three extracellular loops, and three intracellular loops, and a carboxy terminal intracellular domain. Upon binding of a ligand to an extracellular portion of a GPCR, a signal is transduced within the cell that results in a change in a biological or physiological property of the cell. GPCRs, along with G-proteins and effectors (intracellular enzymes and channels modulated by G-proteins), are the components of a modular signaling system that connects the state of intracellular second messengers to extracellular inputs.

GPCR genes and gene-products are potential causative agents of disease (Spiegel *et al., J. Clin. Invest. 92*:1119-1125 (1993); McKusick *et al., J. Med. Genet. 30*:1-26 (1993)). Specific defects in the rhodopsin gene and the V2 vasopressin receptor gene have been shown to cause various forms of retinitis pigmentosum. (Nathans *et al., Annu. Rev. Genet. 26*:403-424(1992)), and nephrogenic diabetes insipidus (Holtzman *et al., Hum. Mol. Genet. 2*:1201-1204 (1993)). These receptors are of critical importance to

both the central nervous system and peripheral physiological processes. Evolutionary analyses suggest that the ancestor of these proteins originally developed in concert with complex body plans and nervous systems.

The GPCR protein superfamily can be divided into five families: Family I, receptors typified by rhodopsin and the β2-adrenergic receptor and currently represented by over 200 unique members (Dohlman *et al.*, *Annu. Rev. Biochem. 60*:653-688 (1991)); Family II, the parathyroid hormone/calcitonin/secretin receptor family (Juppner *et al.*, *Science 254*:1024-1026 (1991); Lin *et al.*, *Science 254*:1022-1024 (1991)); Family III, the metabotropic glutamate receptor family (Nakanishi, *Science 258* 597:603 (1992)); Family IV, the cAMP receptor family, important in the chemotaxis and development of *D. discoideum* (Klein *et al.*, *Science 241*:1467-1472 (1988)); and Family V, the fungal mating pheromone receptors such as STE2 (Kurjan, *Annu. Rev. Biochem. 61*:1097-1129 (1992)).

There are also a small number of other proteins which present seven putative hydrophobic segments and appear to be unrelated to GPCRs; they have not been shown to couple to G-proteins. *Drosophila* expresses a photoreceptor-specific protein, bride of sevenless (boss), a seven-transmembrane-segment protein which has been extensively studied and does not show evidence of being a GPCR (Hart *et al.*, *Proc. Natl. Acad. Sci. USA 90*:5047-5051 (1993)). The gene *frizzled* (*fz*) in *Drosophila* is also thought to be a protein with seven transmembrane segments. Like boss, fz has not been shown to couple to G-proteins (Vinson *et al.*, *Nature 338*:263-264 (1989)).

G proteins represent a family of heterotrimeric proteins composed of α, β and γ subunits, that bind guanine nucleotides. These proteins are usually linked to cell surface receptors, e.g., receptors containing seven transmembrane segments. Following ligand binding to the GPCR, a conformational change is transmitted to the G protein, which causes the α-subunit to exchange a bound GDP molecule for a GTP molecule and to dissociate from the βγ-subunits. The GTP-bound form of the α-subunit typically functions as an effector-modulating moiety, leading to the production of second messengers, such as cAMP (e.g., by activation of adenyl cyclase), diacylglycerol or inositol phosphates. Greater than 20 different types of α-subunits are known in humans. These subunits associate with a smaller pool of β and γ subunits. Examples of mammalian G proteins include Gi, Go, Gq, Gs and Gt. G proteins are described

extensively in Lodish *et al.*, *Molecular Cell Biology*, (Scientific American Books Inc., New York, N.Y., 1995), the contents of which are incorporated herein by reference. GPCRs, G proteins and G protein-linked effector and second messenger systems have been reviewed in *The G-Protein Linked Receptor Fact Book*, Watson *et al.*, eds., Academic Press (1994).

GPCRs are a major target for drug action and development. Accordingly, it is valuable to the field of pharmaceutical development to identify and characterize previously unknown GPCRs or novel seven transmembrane proteins that are putative GPCRs or may be biochemically relevant to GPCR function. The present invention advances the state of the art by providing a previously unidentified human seven transmembrane protein, which may represent a new human GPCR or a molecule related to biochemical functioning of GPCRs.

SUMMARY OF THE INVENTION

It is an object of the invention to identify novel GPCRs.

It is a further object of the invention to provide novel GPCR polypeptides that are useful as reagents or targets in receptor assays applicable to treatment and diagnosis of GPCR-mediated disorders.

It is a further object of the invention to provide polynucleotides corresponding to the novel GPCR receptor polypeptides that are useful as targets and reagents in receptor assays applicable to treatment and diagnosis of GPCR-mediated disorders and useful for producing novel receptor polypeptides by recombinant methods.

A specific object of the invention is to identify compounds that act as agonists and antagonists and modulate the expression of the novel receptor.

A further specific object of the invention is to provide compounds that modulate expression of the receptor for treatment and diagnosis of GPCR- related disorders.

The invention is thus based on the identification of a novel seven transmembrane protein, which may represent a novel human GPCR, designated herein the 19459 protein.

The invention provides isolated 19459 polypeptides including a polypeptide having the amino acid sequence shown in SEQ ID NO:1, or the amino acid sequence

encoded by the cDNA insert of the plasmid deposited with ATCC as Patent Deposit Number _____ on _____ ("the deposited cDNA").

The invention also provides isolated 19459 nucleic acid molecules having the sequence shown in SEQ ID NO:1, SEQ ID NO:3, or in the deposited cDNA.

The invention also provides variant polypeptides having an amino acid sequence that is substantially homologous to the amino acid sequence shown in SEQ ID NO:2 or encoded by the deposited cDNA.

The invention also provides variant nucleic acid sequences that are substantially homologous to the nucleotide sequence shown in SEQ ID NO:1, SEQ ID NO:3, or in the deposited cDNA.

The invention also provides fragments of the polypeptide shown in SEQ ID NO:2 and nucleotide shown in SEQ ID NO:1 or SEQ ID NO:3, as well as substantially homologous fragments of the polypeptide or nucleic acid.

The invention further provides nucleic acid constructs comprising the nucleic acid molecules described above. In a preferred embodiment, the nucleic acid molecules of the invention are operatively linked to a regulatory sequence.

The invention also provides vectors and host cells for expressing the 19459 nucleic acid molecules and polypeptides and particularly recombinant vectors and host cells.

The invention also provides methods of making the vectors and host cells and methods for using them to produce the 19459 nucleic acid molecules and polypeptides.

The invention also provides antibodies or antigen-binding fragments thereof that selectively bind the 19459 polypeptides and fragments.

The invention also provides methods of screening for compounds that modulate expression or activity of the 19459 polypeptides or nucleic acid (RNA or DNA).

The invention also provides a process for modulating 19459 polypeptide or nucleic acid expression or activity, especially using the screened compounds. Modulation may be used to treat conditions related to aberrant activity or expression of the 19459 polypeptides or nucleic acids.

The invention also provides assays for determining the presence or absence of and level of the 19459 polypeptides or nucleic acid molecules in a biological sample, including for disease diagnosis.

The invention also provides assays for determining the presence of a mutation in the 19459 polypeptides or nucleic acid molecules, including for disease diagnosis.

In still a further embodiment, the invention provides a computer readable means containing the nucleotide and/or amino acid sequences of the nucleic acids and polypeptides of the invention, respectively.

DESCRIPTION OF THE DRAWINGS

Figure 1 shows the 19459 nucleotide sequence (SEQ ID NO:1) and the deduced 19459 amino acid sequence (SEQ ID NO:2). The 19459 coding sequence is shown in SEQ ID NO:3

Figure 2 shows a hydrophobicity plot for the 19459 polypeptide. Relative hydrophobic residues are shown above the dashed horizontal line, and relative hydrophilic residues are below the dashed horizontal line. The cysteine residues (cys) and N glycosylation site (Ngly) are indicated by short vertical lines just below the hydropathy trace. The numbers corresponding to the amino acid sequence (shown in SEQ ID NO:2) of human 19459 are indicated . Polypeptides of the invention include fragments which include: all or a part of a hydrophobic sequence (a sequence above the dashed line); or all or part of a hydrophilic fragment (a sequence below the dashed line). Other fragments include a cysteine residue or as N-glycosylation site.

Figure 3 shows an analysis of the 19459 amino acid sequence: αβturn and coil regions; hydrophilicity; amphipathic regions; flexible regions; antigenic index; and surface probability plot.

Figure 4 shows an analysis of the 19459 open reading frame for amino acids corresponding to specific functional sites. Glycosylation sites are found from about amino acid 4 to about amino acid 7, from about amino acid 236 to about amino acid 239, and from about amino acid 257 to about amino acid 260. A cAMP- and cGMP-dependent protein kinase phosphorylation site is found from about amino acid 54 to about amino acid 57. Protein kinase C phosphorylation sites are found from about amino acid 84 to about amino acid 86, and from about amino acid 161 to about amino acid 163. Casein kinase II phosphorylation sites are found from about amino acid 246 to about amino acid 249, and from about amino acid 338 to about amino acid 341. N-myristoylation sites are found from about amino acid 7 to about amino acid 12, from

about amino acid 13 to about amino acid 18, from about amino acid 25 to about amino acid 30, from amino acid 141 to about amino acid 146, from about amino acid 212 to about amino acid 217, from about amino acid 312 to about amino acid 317, from about amino acid 347 to about amino acid 352, and from about amino acid 366 to about amino acid 371. In the case of glycosylation, the actual modified residue is the first amino acid. In the case of cAMP- and cGMP-dependent protein kinase phosphorylation, the actual modified residue is the last amino acid. In the case of protein kinase C phosphorylation, casein kinase II phosphorylation, and N-myristoylation, the actual modified residue is the first amino acid. Predicted transmembrane segments for the presumed mature peptide are also included in this figure.

Figure 5 shows relative expression of the 19459 protein in various normal human tissues and in diseased heart tissues. Expression levels were determined by quantitative RT (reverse transcriptase PCR (Taqman® brand PCR kit, Applied Biosystems). The quantitative RT-PCR reactions were performed according to the kit manufacturer's instructions.

Figure 6 shows expression of 39406 in the following human tissues and cell types: lung (column 1), kidney (column 2), brain (column 3), hear (column 4), colon (column 5), tonsil (column 6), spleen (column 7), fetal liver (column 8), a pool of seven normal livers (column 9), resting stellate cells (column 10), serum-reactivated stellate cells (column 11), normal human liver lung fibroblasts (column 12), normal human liver lung fibroblasts treated with TGF-β for 48 hours (column 13), HepG2 cells (column 14), HepG2 cells treated with TGF-β for 48 hours (column 15), resting NHLH cells (column 16), activated normal human lung fibroblasts (column 17), tissue from fibrotic liver (columns 18-21), Th1 cells (columns 22 and 24), Th2 cells (columns 23 and 25), granulocytes (column 26), CD19 positive cells (column 27), CD14 positive cells (column 28), activated CD14 positive cells (column 29), peripheral blood mononuclear cells (column 30), peripheral blood mononuclear cells treated with PHA (column 31), IL-10 (column 32), or IL-13 (column 33); normal human bronchial epithelial cells (column 34), normal human bronchial epithelial cells treated with IL-13 (column 35), Th0 cells (column 36), Th2 cells (column 37), bone marrow mononuclear cells (column 38), mobilized peripheral blood CD34 positive cells (column 39), adult bone marrow CD34 positive cells (column 40), mobilized

bone marrow CD34 positive cells (column 41), erythroid cells (column 42), megakaryocytes (column 43), neutrophils (column 44), mobilized bone marrow CD11b positive cells (column 45), mobilized bone marrow CD15 positive cells (column 46), mobilized bone marrow CD11b negative cells (column 47), glycophroin A positive bone marrow cells (column 48), CD71 positive bone marrow cells (column 49), HepG2A cells (column 50), and HepG2A cells infected with hepatitis B virus (column 51). Expression levels were determined by quantitative RT-PCR as described in the legend for Figure 5.

Figure 7 shows the expression of 19459 in normal human tissues. Expression levels were determined by quantitative RT-PCR as described in the legend for Figure 5.

DETAILED DESCRIPTION OF THE INVENTION

Receptor function/signal pathway

The 19459 receptor protein shares sequence homology with the GPCR family of proteins, which participate in signaling pathways. As used herein, a "signaling pathway" refers to the modulation (e.g., stimulation or inhibition) of a cellular function/activity upon the binding of a ligand to the GPCR (19459 protein). Examples of such functions include mobilization of intracellular molecules that participate in a signal transduction pathway, e.g., phosphatidylinositol 4,5-bisphosphate (PIP_2), inositol 1,4,5-triphosphate (IP_3) and adenylate cyclase; polarization of the plasma membrane; production or secretion of molecules; alteration in the structure of a cellular component; cell proliferation, e.g., synthesis of DNA; cell migration; cell differentiation; and cell survival. Since the 19459 receptor protein is expressed in heart, kidney, muscle, liver, placenta, and tonsils, lung, brain, colon, spleen, liver, granulocytes, bronchial epithelial cells, , Th0 cells, CD34 positive cells from mobilized peripheral blood, adult bone marrow, and mobilized bone marrow, neutrophils, osteoclasts, thyroid, and testis, cells participating in a 19459 receptor protein signaling pathway may include, but are not limited to cells derived from these tissues.

The response mediated by the receptor protein depends on the type of cell. For example, in some cells, binding of a ligand to the receptor protein may stimulate an activity such as release of compounds, gating of a channel, cellular adhesion, migration,

differentiation, etc., through phosphatidylinositol or cyclic AMP metabolism and turnover while in other cells, the binding of the ligand will produce a different result. Regardless of the cellular activity/response modulated by the receptor protein, the protein, as a GPCR, would interact with G proteins to produce one or more secondary signals, in a variety of intracellular signal transduction pathways, e.g., through phosphatidylinositol or cyclic AMP metabolism and turnover, in a cell.

As used herein, "phosphatidylinositol turnover and metabolism" refers to the molecules involved in the turnover and metabolism of phosphatidylinositol 4,5-bisphosphate (PIP_2) as well as to the activities of these molecules. PIP_2 is a phospholipid found in the cytosolic leaflet of the plasma membrane. Binding of ligand to the receptor activates, in some cells, the plasma-membrane enzyme phospholipase C that in turn can hydrolyze PIP_2 to produce 1,2-diacylglycerol (DAG) and inositol 1,4,5-triphosphate (IP_3). Once formed IP_3 can diffuse to the endoplasmic reticulum surface where it can bind an IP_3 receptor, e.g., a calcium channel protein containing an IP_3 binding site. IP_3 binding can induce opening of the channel, allowing calcium ions to be released into the cytoplasm. IP_3 can also be phosphorylated by a specific kinase to form inositol 1,3,4,5-tetraphosphate (IP_4), a molecule which can cause calcium entry into the cytoplasm from the extracellular medium. IP_3 and IP_4 can subsequently be hydrolyzed very rapidly to the inactive products inositol 1,4-biphosphate (IP_2) and inositol 1,3,4-triphosphate, respectively. These inactive products can be recycled by the cell to synthesize PIP_2. The other second messenger produced by the hydrolysis of PIP_2, namely 1,2-diacylglycerol (DAG), remains in the cell membrane where it can serve to activate the enzyme protein kinase C. Protein kinase C is usually found soluble in the cytoplasm of the cell, but upon an increase in the intracellular calcium concentration, this enzyme can move to the plasma membrane where it can be activated by DAG. The activation of protein kinase C in different cells results in various cellular responses such as the phosphorylation of glycogen synthase, or the phosphorylation of various transcription factors, e.g., NF-kB. The language "phosphatidylinositol activity", as used herein, refers to an activity of PIP_2 or one of its metabolites.

Another signaling pathway in which the receptor may participate is the cAMP turnover pathway. As used herein, "cyclic AMP turnover and metabolism" refers to the molecules involved in the turnover and metabolism of cyclic AMP (cAMP) as well as to the activities of these molecules. Cyclic AMP is a second messenger produced in

response to ligand-induced stimulation of certain G protein coupled receptors. In the cAMP signaling pathway, binding of a ligand to a GPCR can lead to the activation of the enzyme adenyl cyclase, which catalyzes the synthesis of cAMP. The newly synthesized cAMP can in turn activate a cAMP-dependent protein kinase. This activated kinase can phosphorylate a voltage-gated potassium channel protein, or an associated protein, and lead to the inability of the potassium channel to open during an action potential. The inability of the potassium channel to open results in a decrease in the outward flow of potassium, which normally repolarizes the membrane of a neuron, leading to prolonged membrane depolarization.

Polypeptides

The invention is based on the identification of a novel human seven transmembrane protein, potentially a novel human G-coupled protein receptor. Specifically, an expressed sequence tag (EST) was selected based on homology to G-protein-coupled receptor sequences. This EST was used to design primers based on primary sequences that it contains and used to identify a cDNA from a human cDNA library. Positive clones were sequenced and the overlapping fragments were assembled. Analysis of the assembled sequence revealed that the cloned cDNA molecule encodes a seven transmembrane protein, potentially a G-protein coupled receptor.

The invention thus relates to a novel seven transmembrane protein having the deduced amino acid sequence shown in Figure 1 (SEQ ID NO:2) or having the amino acid sequence encoded by cDNA insert of the plasmid deposited with American Type Culture Collection (ATCC), 10801 University Boulevard, Manassas, VA 20110-2209 on as Patent Deposit No _____.

The deposit will be maintained under the terms of the Budapest Treaty on the International Recognition of the Deposit of Microorganisms. The deposit is provided as a convenience to those of skill in the art and is not an admission that a deposit is required under 35 U.S.C. § 112. The deposited sequence, as well as the polypeptide encoded by the sequence, is incorporated herein by reference and controls in the event of any conflict, such as a sequencing error, with description in this application.

The "19459 polypeptide" or "19459 protein" refers to the polypeptide in SEQ ID NO:2 or the polypeptide encoded by the cDNA insert of the deposited plasmid. The

terms, however, further include the numerous variants described herein, as well as fragments derived from the full length 19459 polypeptide and variants.

The present invention thus provides an isolated or purified 19459 polypeptide and variants and fragments thereof.

The 19459 polypeptide is a 397 residue protein exhibiting three main structural domains, an amino terminal extracellular domain, a transmembrane domain, and a carboxy terminal intracellular domain. The transmembrane domain contains seven segments that span the membrane. Within the region spanning the entire transmembrane domain are three intracellular and three extracellular loops.

Based on a BLAST search, homology was shown to a human seven transmembrane protein.

19459 nucleic acid is highly expressed in tissues or cells that include, but are not limited to, those shown in Figure 5, 6, and 7.

As used herein, a polypeptide is said to be "isolated" or "purified" when it is substantially free of cellular material when it is isolated from recombinant and non-recombinant cells, or free of chemical precursors or other chemicals when it is chemically synthesized. A polypeptide, however, can be joined to another polypeptide with which it is not normally associated in a cell and still be considered "isolated" or "purified."

The 19459 polypeptides can be purified to homogeneity. It is understood, however, that preparations in which the polypeptide is not purified to homogeneity are useful and considered to contain an isolated form of the polypeptide. The critical feature is that the preparation allows for the desired function of the polypeptide, even in the presence of considerable amounts of other components. Thus, the invention encompasses various degrees of purity.

In one embodiment, the language "substantially free of cellular material" includes preparations of the 19459 polypeptide having less than about 30% (by dry weight) other proteins (i.e., contaminating protein), less than about 20% other proteins, less than about 10% other proteins, or less than about 5% other proteins. When the polypeptide is recombinantly produced, it can also be substantially free of culture medium, i.e., culture medium represents less than about 20%, less than about 10%, or less than about 5% of the volume of the protein preparation.

A polypeptide is also considered to be isolated when it is part of a membrane preparation or is purified and then reconstituted with membrane vesicles or liposomes.

The language "substantially free of chemical precursors or other chemicals" includes preparations of the polypeptide in which it is separated from chemical precursors or other chemicals that are involved in its synthesis. In one embodiment, the language "substantially free of chemical precursors or other chemicals" includes preparations of the polypeptide having less than about 30% (by dry weight) chemical precursors or other chemicals, less than about 20% chemical precursors or other chemicals, less than about 10% chemical precursors or other chemicals, or less than about 5% chemical precursors or other chemicals.

In one embodiment, the polypeptide comprises the amino acid sequence shown in SEQ ID NO:2. However, the invention also encompasses sequence variants. By "variants" is intended proteins or polypeptides having an amino acid sequence that is at least about 60%, preferably about 65%, 75%, 85%, 95%, 96%, 97%, 98%, or 99% identical to the amino acid sequence of SEQ ID NO:2. Variants also include polypeptides encoded by the cDNA insert of the plasmid deposited with ATCC as Patent Deposit No._____, or polypeptides encoded by a nucleic acid molecule that hybridizes to the complement of the nucleic acid molecule of SEQ ID NO:1 or SEQ ID NO:3 under stringent conditions. In another embodiment, a variant of an isolated polypeptide of the present invention differs, by at least 1, but less than 5, 10, 20, 50, or 100 amino acid residues from the sequence shown in SEQ ID NO:2. If alignment is needed for this comparison the sequences should be aligned for maximum identity. "Looped" out sequences from deletions or insertions, or mismatches, are considered differences. Such variants generally retain the functional (biological) activity of the 19459 proteins of the invention. Variants include polypeptides that differ in amino acid sequence due to natural allelic variation or mutagenesis. Variants also encompass proteins derived from other genetic loci in an organism, but having substantial homology to the 19459 protein of SEQ ID NO:2. Variants also include proteins substantially homologous to the 19459 protein but derived from another organism, i.e., an ortholog. Variants also include proteins that are substantially homologous to the 19459 protein that are produced by chemical synthesis. Variants also include proteins that are substantially homologous to the 19459 protein that are produced by recombinant methods. It is understood, however, that variants exclude any amino acid sequences disclosed prior to the invention.

To determine the percent identity of two amino acid sequences, or of two nucleic acid sequences, the sequences are aligned for optimal comparison purposes (e.g., gaps can be introduced in one or both of a first and a second amino acid or nucleic acid sequence for optimal alignment and non-homologous sequences can be disregarded for comparison purposes). In a preferred embodiment, the length of a reference sequence aligned for comparison purposes is at least 30%, preferably at least 40%, more preferably at least 50%, even more preferably at least 60%, and even more preferably at least 70%, 80%, 90%, 100% of the length of the reference sequence. The amino acid residues or nucleotides at corresponding amino acid positions or nucleotide positions are then compared. When a position in the first sequence is occupied by the same amino acid residue or nucleotide as the corresponding position in the second sequence, then the molecules are identical at that position (as used herein amino acid or nucleic acid "identity" is equivalent to amino acid or nucleic acid "homology"). The percent identity between the two sequences is a function of the number of identical positions shared by the sequences, taking into account the number of gaps, and the length of each gap, which need to be introduced for optimal alignment of the two sequences.

The comparison of sequences and determination of percent identity between two sequences can be accomplished using a mathematical algorithm. In a preferred embodiment, the percent identity between two amino acid sequences is determined using the Needleman and Wunsch (1970) J. Mol. Biol. 48:444-453 algorithm which has been incorporated into the GAP program in the GCG software package (available at http://www.gcg.com), using either a Blossum 62 matrix or a PAM250 matrix, and a gap weight of 16, 14, 12, 10, 8, 6, or 4 and a length weight of 1, 2, 3, 4, 5, or 6. In yet another preferred embodiment, the percent identity between two nucleotide sequences is determined using the GAP program in the GCG software package (available at http://www.gcg.com), using a NWSgapdna.CMP matrix and a gap weight of 40, 50, 60, 70, or 80 and a length weight of 1, 2, 3, 4, 5, or 6. A particularly preferred set of parameters (and the one that should be used if the practitioner is uncertain about what parameters should be applied to determine if a molecule is within a sequence identity or homology limitation of the invention) is using a Blossum 62 scoring matrix with a gap open penalty of 12, a gap extend penalty of 4, and a frameshift gap penalty of 5.

The percent identity between two amino acid or nucleotide sequences can be determined using the algorithm of E. Meyers and W. Miller (1989) CABIOS 4:11-17 which has been incorporated into the ALIGN program (version 2.0), using a PAM120 weight residue table, a gap length penalty of 12 and a gap penalty of 4.

The nucleic acid and protein sequences described herein can be used as a "query sequence" to perform a search against public databases to, for example, identify other family members or related sequences. Such searches can be performed using the NBLAST and XBLAST programs (version 2.0) of Altschul, et al. (1990) J. Mol. Biol. 215:403-10. BLAST nucleotide searches can be performed with the NBLAST program, score = 100, wordlength = 12 to obtain nucleotide sequences homologous to 19459 nucleic acid molecules of the invention. BLAST protein searches can be performed with the XBLAST program, score = 50, wordlength = 3 to obtain amino acid sequences homologous to 19459 protein molecules of the invention. To obtain gapped alignments for comparison purposes, Gapped BLAST can be utilized as described in Altschul et al. (1997) Nucleic Acids Res. 25(17):3389-3402. When utilizing BLAST and Gapped BLAST programs, the default parameters of the respective programs (e.g., XBLAST and NBLAST) can be used. See http://www.ncbi.nlm.nih.gov.

The invention also encompasses polypeptides having a lower degree of identity but having sufficient similarity so as to perform one or more of the same functions performed by the 19459 polypeptide. Similarity is determined by conserved amino acid substitution. Such substitutions are those that substitute a given amino acid in a polypeptide by another amino acid of like characteristics. Conservative substitutions are likely to be phenotypically silent. Typically seen as conservative substitutions are the replacements, one for another, among the aliphatic amino acids Ala, Val, Leu, and Ile; interchange of the hydroxyl residues Ser and Thr, exchange of the acidic residues Asp and Glu, substitution between the amide residues Asn and Gln, exchange of the basic residues Lys and Arg and replacements among the aromatic residues Phe, Tyr. Guidance concerning which amino acid changes are likely to be phenotypically silent are found in Bowie *et al.*, *Science 247*:1306-1310 (1990).

TABLE 1. Conservative Amino Acid Substitutions.

Aromatic	Phenylalanine
	Tryptophan
	Tyrosine
Hydrophobic	Leucine
	Isoleucine
	Valine
Polar	Glutamine
	Asparagine
Basic	Arginine
	Lysine
	Histidine
Acidic	Aspartic Acid
	Glutamic Acid
Small	Alanine
	Serine
	Threonine
	Methionine
	Glycine

A variant polypeptide can differ in amino acid sequence by one or more substitutions, deletions, insertions, inversions, fusions, and truncations or a combination of any of these.

Variant polypeptides can be fully functional or can lack function in one or more activities. Thus, in the present case, variations can affect the function, for example, of one or more regions corresponding to ligand binding, membrane association, G-protein binding and signal transduction.

Fully functional variants typically contain only conservative variation or variation in non-critical residues or in non-critical regions. Functional variants can also

contain substitution of similar amino acids which result in no change or an insignificant change in function. Alternatively, such substitutions may positively or negatively affect function to some degree.

Non-functional variants typically contain one or more non-conservative amino acid substitutions, deletions, insertions, inversions, or truncation or a substitution, insertion, inversion, or deletion in a critical residue or critical region.

As indicated, variants can be naturally-occurring or can be made by recombinant means or chemical synthesis to provide useful and novel characteristics for the 19459 polypeptide. This includes preventing immunogenicity from pharmaceutical formulations by preventing protein aggregation.

Useful variations further include alteration of ligand binding characteristics. For example, one embodiment involves a variation at the binding site that results in binding but not release, or slower release, of ligand. A further useful variation at the same sites can result in a higher affinity for ligand. Useful variations also include changes that provide for affinity for another ligand. Another useful variation includes one that allows binding but which prevents activation by the ligand. Another useful variation includes variation in the transmembrane G-protein-binding/signal transduction domain that provides for reduced or increased binding by the appropriate G-protein or for binding by a different G-protein than the one with which the receptor is normally associated. Another useful variation provides a fusion protein in which one or more domains or subregions is operationally fused to one or more domains or subregions from another G-protein coupled receptor or seven transmembrane protein.

Amino acids that are essential for function can be identified by methods known in the art, such as site-directed mutagenesis or alanine-scanning mutagenesis (Cunningham *et al.*, *Science 244*:1081-1085 (1989)). The latter procedure introduces single alanine mutations at every residue in the molecule. The resulting mutant molecules are then tested for biological activity such as receptor binding or *in vitro*, or *in vivo* proliferative activity. Sites that are critical for ligand-receptor binding can also be determined by structural analysis such as crystallization, nuclear magnetic resonance or photoaffinity labeling (Smith *et al.*, *J. Mol. Biol. 224*:899-904 (1992); de Vos *et al.* *Science 255*:306-312 (1992)).

Substantial homology can be to the entire nucleic acid or amino acid sequence or to fragments of these sequences.

The invention thus also includes polypeptide fragments of the 19459 protein. Fragments can be derived from the amino acid sequence shown in SEQ ID NO:2. However, the invention also encompasses fragments of the variants of the 19459 protein as described herein.

The fragments to which the invention pertains, however, are not to be construed as encompassing fragments that may be disclosed prior to the present invention.

As used herein, a fragment comprises at least 10-20, 20-25, 25-30, or more contiguous amino acids. Fragments can retain one or more of the biological activities of the protein, for example the ability to bind to a G-protein or ligand, as well as fragments that can be used as an immunogen to generate antibodies.

Biologically active fragments (peptides which are, for example, around 5-10, 10-15, 15-20, 30, 35, 36, 37, 38, 39, 40, 50, 100, 150, 200, 250, 300, 350, or more amino acids in length) can comprise a domain or motif, e.g., an extracellular or intracellular domain or loop, one or more transmembrane segments, or parts thereof, G-protein binding site, or glycosylation sites, phosphorylation sites, and myristoylation sites. Such domains or motifs can be identified by means of routine computerized homology searching procedures.

Possible fragments include, but are not limited to: 1) soluble peptides comprising the entire amino terminal extracellular domain; 2) peptides comprising the entire carboxy terminal intracellular domain or parts thereof; 3) peptides comprising the region spanning the entire transmembrane domain or parts thereof; 4) any of the specific transmembrane segments, or parts thereof; 5) any of the three intracellular or three extracellular loops, or parts thereof.

Fragments further include combinations of the above fragments, such as an amino terminal domain combined with one or more transmembrane segments and the attendant extra or intracellular loops or one or more transmembrane segments, and the attendant intra or extracellular loops, plus the carboxy terminal domain. Thus, any of the above fragments can be combined. Other fragments include the mature protein from about amino acid 44 to 397. Other fragments contain the various functional sites described herein. Fragments, for example, can extend in one or both directions from the functional site to encompass 5, 10, 15, 20, 30, 40, 50, or up to 100 amino acids. Further, fragments can include sub-fragments of the specific domains mentioned above, which sub-fragments retain the function of the domain from which they are derived.

These regions can be identified by well-known methods involving computerized homology analysis.

Fragments also include antigenic fragments and specifically those shown to have a high antigenicity index in Figure 3.

Further possible fragments include but are not limited to fragments defining a ligand-binding site, fragments defining membrane association, fragments defining interaction with G proteins and signal transduction. By this is intended a discrete fragment that provides the relevant function or allows the relevant function to be identified. In a preferred embodiment, the fragment contains a ligand-binding site.

The invention also provides fragments with immunogenic properties. These contain an epitope-bearing portion of the 19459 protein and variants. These epitope-bearing peptides are useful to raise antibodies that bind specifically to a 19459 polypeptide or region or fragment. These peptides can contain at least 6, 10, 12, at least 14, or between at least about 15 to about 30 amino acids.

Non-limiting examples of antigenic polypeptides that can be used to generate antibodies include peptides derived from the amino terminal extracellular domain or any of the extracellular loops. Regions having a high antigenicity index are shown in Figure 3. However, intracellularly-made antibodies ("intrabodies") are also encompassed, which would recognize intracellular peptide regions.

The 19459 polypeptides (including variants and fragments which may have been disclosed prior to the present invention) are useful for biological assays related to seven transmembrane protein and especially GPCRs. Such assays involve any of the known seven transmembrane protein or GPCR functions or activities or properties useful for diagnosis and treatment of seven transmembrane protein-related and especially GPCR-related conditions, especially diseases involving the tissues in which the protein is expressed as disclosed herein.

Disorders involving the lung include, but are not limited to, congenital anomalies; atelectasis; diseases of vascular origin, such as pulmonary congestion and edema, including hemodynamic pulmonary edema and edema caused by microvascular injury, adult respiratory distress syndrome (diffuse alveolar damage), pulmonary embolism, hemorrhage, and infarction, and pulmonary hypertension and vascular sclerosis; chronic obstructive pulmonary disease, such as emphysema, chronic bronchitis, bronchial asthma, and bronchiectasis; diffuse interstitial

(infiltrative, restrictive) diseases, such as pneumoconioses, sarcoidosis, idiopathic pulmonary fibrosis, desquamative interstitial pneumonitis, hypersensitivity pneumonitis, pulmonary eosinophilia (pulmonary infiltration with eosinophilia), *Bronchiolitis obliterans*-organizing pneumonia, diffuse pulmonary hemorrhage syndromes, including Goodpasture syndrome, idiopathic pulmonary hemosiderosis and other hemorrhagic syndromes, pulmonary involvement in collagen vascular disorders, and pulmonary alveolar proteinosis; complications of therapies, such as drug-induced lung disease, radiation-induced lung disease, and lung transplantation; tumors, such as bronchogenic carcinoma, including paraneoplastic syndromes, bronchioloalveolar carcinoma, neuroendocrine tumors, such as bronchial carcinoid, miscellaneous tumors, and metastatic tumors; pathologies of the pleura, including inflammatory pleural effusions, noninflammatory pleural effusions, pneumothorax, and pleural tumors, including solitary fibrous tumors (pleural fibroma) and malignant mesothelioma.

Disorders involving the liver include, but are not limited to, hepatic injury; jaundice and cholestasis, such as bilirubin and bile formation; hepatic failure and cirrhosis, such as cirrhosis, portal hypertension, including ascites, portosystemic shunts, and splenomegaly; infectious disorders, such as viral hepatitis, including hepatitis A-E infection and infection by other hepatitis viruses, clinicopathologic syndromes, such as the carrier state, asymptomatic infection, acute viral hepatitis, chronic viral hepatitis, and fulminant hepatitis; autoimmune hepatitis; drug- and toxin-induced liver disease, such as alcoholic liver disease; inborn errors of metabolism and pediatric liver disease, such as hemochromatosis, Wilson disease, a_1-antitrypsin deficiency, and neonatal hepatitis; intrahepatic biliary tract disease, such as secondary biliary cirrhosis, primary biliary cirrhosis, primary sclerosing cholangitis, and anomalies of the biliary tree; circulatory disorders, such as impaired blood flow into the liver, including hepatic artery compromise and portal vein obstruction and thrombosis, impaired blood flow through the liver, including passive congestion and centrilobular necrosis and peliosis hepatis, hepatic vein outflow obstruction, including hepatic vein thrombosis (Budd-Chiari syndrome) and veno-occlusive disease; hepatic disease associated with pregnancy, such as preeclampsia and eclampsia, acute fatty liver of pregnancy, and intrehepatic cholestasis of pregnancy; hepatic complications of organ or bone marrow transplantation, such as drug toxicity after bone marrow

transplantation, graft-versus-host disease and liver rejection, and nonimmunologic damage to liver allografts; tumors and tumorous conditions, such as nodular hyperplasias, adenomas, and malignant tumors, including primary carcinoma of the liver and metastatic tumors.

Disorders involving the kidney include, but are not limited to, congenital anomalies including, but not limited to, cystic diseases of the kidney, that include but are not limited to, cystic renal dysplasia, autosomal dominant (adult) polycystic kidney disease, autosomal recessive (childhood) polycystic kidney disease, and cystic diseases of renal medulla, which include, but are not limited to, medullary sponge kidney, and nephronophthisis-uremic medullary cystic disease complex, acquired (dialysis-associated) cystic disease, such as simple cysts; glomerular diseases including pathologies of glomerular injury that include, but are not limited to, in situ immune complex deposition, that includes, but is not limited to, anti-GBM nephritis, Heymann nephritis, and antibodies against planted antigens, circulating immune complex nephritis, antibodies to glomerular cells, cell-mediated immunity in glomerulonephritis, activation of alternative complement pathway, epithelial cell injury, and pathologies involving mediators of glomerular injury including cellular and soluble mediators, acute glomerulonephritis, such as acute proliferative (poststreptococcal, postinfectious) glomerulonephritis, including but not limited to, poststreptococcal glomerulonephritis and nonstreptococcal acute glomerulonephritis, rapidly progressive (crescentic) glomerulonephritis, nephrotic syndrome, membranous glomerulonephritis (membranous nephropathy), minimal change disease (lipoid nephrosis), focal segmental glomerulosclerosis, membranoproliferative glomerulonephritis, IgA nephropathy (Berger disease), focal proliferative and necrotizing glomerulonephritis (focal glomerulonephritis), hereditary nephritis, including but not limited to, Alport syndrome and thin membrane disease (benign familial hematuria), chronic glomerulonephritis, glomerular lesions associated with systemic disease, including but not limited to, systemic lupus erythematosus, Henoch-Schönlein purpura, bacterial endocarditis, diabetic glomerulosclerosis, amyloidosis, fibrillary and immunotactoid glomerulonephritis, and other systemic disorders; diseases affecting tubules and interstitium, including acute tubular necrosis and tubulointerstitial nephritis, including but not limited to, pyelonephritis and urinary tract infection, acute pyelonephritis, chronic pyelonephritis and reflux nephropathy, and tubulointerstitial nephritis induced

by drugs and toxins, including but not limited to, acute drug-induced interstitial nephritis, analgesic abuse nephropathy, nephropathy associated with nonsteroidal anti-inflammatory drugs, and other tubulointerstitial diseases including, but not limited to, urate nephropathy, hypercalcemia and nephrocalcinosis, and multiple myeloma; diseases of blood vessels including benign nephrosclerosis, malignant hypertension and accelerated nephrosclerosis, renal artery stenosis, and thrombotic microangiopathies including, but not limited to, classic (childhood) hemolytic-uremic syndrome, adult hemolytic-uremic syndrome/thrombotic thrombocytopenic purpura, idiopathic HUS/TTP, and other vascular disorders including, but not limited to, atherosclerotic ischemic renal disease, atheroembolic renal disease, sickle cell disease nephropathy, diffuse cortical necrosis, and renal infarcts; urinary tract obstruction (obstructive uropathy); urolithiasis (renal calculi, stones); and tumors of the kidney including, but not limited to, benign tumors, such as renal papillary adenoma, renal fibroma or hamartoma (renomedullary interstitial cell tumor), angiomyolipoma, and oncocytoma, and malignant tumors, including renal cell carcinoma (hypernephroma, adenocarcinoma of kidney), which includes urothelial carcinomas of renal pelvis.

Disorders involving the skeletal muscle include tumors such as rhabdomyosarcoma.

Disorders involving the brain include, but are not limited to, disorders involving neurons, and disorders involving glia, such as astrocytes, oligodendrocytes, ependymal cells, and microglia; cerebral edema, raised intracranial pressure and herniation, and hydrocephalus; malformations and developmental diseases, such as neural tube defects, forebrain anomalies, posterior fossa anomalies, and syringomyelia and hydromyelia; perinatal brain injury; cerebrovascular diseases, such as those related to hypoxia, ischemia, and infarction, including hypotension, hypoperfusion, and low-flow states--global cerebral ischemia and focal cerebral ischemia--infarction from obstruction of local blood supply, intracranial hemorrhage, including intracerebral (intraparenchymal) hemorrhage, subarachnoid hemorrhage and ruptured berry aneurysms, and vascular malformations, hypertensive cerebrovascular disease, including lacunar infarcts, slit hemorrhages, and hypertensive encephalopathy; infections, such as acute meningitis, including acute pyogenic (bacterial) meningitis and acute aseptic (viral) meningitis, acute focal suppurative infections, including brain abscess, subdural empyema, and extradural abscess, chronic bacterial

meningoencephalitis, including tuberculosis and mycobacterioses, neurosyphilis, and neuroborreliosis (Lyme disease), viral meningoencephalitis, including arthropod-borne (Arbo) viral encephalitis, *Herpes simplex* virus Type 1, *Herpes simplex* virus Type 2, *Varicalla-zoster* virus (*Herpes zoster*), cytomegalovirus, poliomyelitis, rabies, and human immunodeficiency virus 1, including HIV-1 meningoencephalitis (subacute encephalitis), vacuolar myelopathy, AIDS-associated myopathy, peripheral neuropathy, and AIDS in children, progressive multifocal leukoencephalopathy, subacute sclerosing panencephalitis, fungal meningoencephalitis, other infectious diseases of the nervous system; transmissible spongiform encephalopathies (prion diseases); demyelinating diseases, including multiple sclerosis, multiple sclerosis variants, acute disseminated encephalomyelitis and acute necrotizing hemorrhagic encephalomyelitis, and other diseases with demyelination; degenerative diseases, such as degenerative diseases affecting the cerebral cortex, including Alzheimer disease and Pick disease, degenerative diseases of basal ganglia and brain stem, including Parkinsonism, idiopathic Parkinson disease (paralysis agitans), progressive supranuclear palsy, corticobasal degeneration, multiple system atrophy, including striatonigral degeneration, Shy-Drager syndrome, and olivopontocerebellar atrophy, and Huntington disease; spinocerebellar degenerations, including spinocerebellar ataxias, including Friedreich ataxia, and ataxia-telanglectasia, degenerative diseases affecting motor neurons, including amyotrophic lateral sclerosis (motor neuron disease), bulbospinal atrophy (Kennedy syndrome), and spinal muscular atrophy; inborn errors of metabolism, such as leukodystrophies, including Krabbe disease, metachromatic leukodystrophy, adrenoleukodystrophy, Pelizaeus-Merzbacher disease, and Canavan disease, mitochondrial encephalomyopathies, including Leigh disease and other mitochondrial encephalomyopathies; toxic and acquired metabolic diseases, including vitamin deficiencies such as thiamine (vitamin B_1) deficiency and vitamin B_{12} deficiency, neurologic sequelae of metabolic disturbances, including hypoglycemia, hyperglycemia, and hepatic encephatopathy, toxic disorders, including carbon monoxide, methanol, ethanol, and radiation, including combined methotrexate and radiation-induced injury; tumors, such as gliomas, including astrocytoma, including fibrillary (diffuse) astrocytoma and glioblastoma multiforme, pilocytic astrocytoma, pleomorphic xanthoastrocytoma, and brain stem glioma, oligodendroglioma, and ependymoma and related paraventricular mass lesions,

neuronal tumors, poorly differentiated neoplasms, including medulloblastoma, other parenchymal tumors, including primary brain lymphoma, germ cell tumors, and pineal parenchymal tumors, meningiomas, metastatic tumors, paraneoplastic syndromes, peripheral nerve sheath tumors, including schwannoma, neurofibroma, and malignant peripheral nerve sheath tumor (malignant schwannoma), and neurocutaneous syndromes (phakomatoses), including neurofibromotosis, including Type 1 neurofibromatosis (NF1) and TYPE 2 neurofibromatosis (NF2), tuberous sclerosis, and Von Hippel-Lindau disease.

Disorders involving T-cells include, but are not limited to, cell-mediated hypersensitivity, such as delayed type hypersensitivity and T-cell-mediated cytotoxicity, and transplant rejection; autoimmune diseases, such as systemic lupus erythematosus, Sjögren syndrome, systemic sclerosis, inflammatory myopathies, mixed connective tissue disease, and polyarteritis nodosa and other vasculitides; immunologic deficiency syndromes, including but not limited to, primary immunodeficiencies, such as thymic hypoplasia, severe combined immunodeficiency diseases, and AIDS; leukopenia; reactive (inflammatory) proliferations of white cells, including but not limited to, leukocytosis, acute nonspecific lymphadenitis, and chronic nonspecific lymphadenitis; neoplastic proliferations of white cells, including but not limited to lymphoid neoplasms, such as precursor T-cell neoplasms, such as acute lymphoblastic leukemia/lymphoma, peripheral T-cell and natural killer cell neoplasms that include peripheral T-cell lymphoma, unspecified, adult T-cell leukemia/lymphoma, mycosis fungoides and Sézary syndrome, and Hodgkin disease.

In normal bone marrow, the myelocytic series (polymorphoneuclear cells) make up approximately 60% of the cellular elements, and the erythrocytic series, 20-30%. Lymphocytes, monocytes, reticular cells, plasma cells and megakaryocytes together constitute 10-20%. Lymphocytes make up 5-15% of normal adult marrow. In the bone marrow, cell types are add mixed so that precursors of red blood cells (erythroblasts), macrophages (monoblasts), platelets (megakaryocytes), polymorphoneuclear leucocytes (myeloblasts), and lymphocytes (lymphoblasts) can be visible in one microscopic field. In addition, stem cells exist for the different cell lineages, as well as a precursor stem cell for the committed progenitor cells of the different lineages. The various types of cells and stages of each would be known to the person of ordinary skill in the art and are found, for example, on page 42 (Figure

2-8) of *Immunology, Imunopathology and Immunity*, Fifth Edition, Sell *et al.* Simon and Schuster (1996), incorporated by reference for its teaching of cell types found in the bone marrow. According, the invention is directed to disorders arising from these cells. These disorders include but are not limited to the following: diseases involving hematopoeitic stem cells; committed lymphoid progenitor cells; lymphoid cells including B and T-cells; committed myeloid progenitors, including monocytes, granulocytes, and megakaryocytes; and committed erythroid progenitors. These include but are not limited to the leukemias, including B-lymphoid leukemias, T-lymphoid leukemias, undifferentiated leukemias; erythroleukemia, megakaryoblastic leukemia, monocytic; [leukemias are encompassed with and without differentiation]; chronic and acute lymphoblastic leukemia, chronic and acute lymphocytic leukemia, chronic and acute myelogenous leukemia, lymphoma, myelo dysplastic syndrome, chronic and acute myeloid leukemia, myelomonocytic leukemia; chronic and acute myeloblastic leukemia, chronic and acute myelogenous leukemia, chronic and acute promyelocytic leukemia, chronic and acute myelocytic leukemia, hematologic malignancies of monocyte-macrophage lineage, such as juvenile chronic myelogenous leukemia; secondary AML, antecedent hematological disorder; refractory anemia; aplastic anemia; reactive cutaneous angioendotheliomatosis; fibrosing disorders involving altered expression in dendritic cells, disorders including systemic sclerosis, E-M syndrome, epidemic toxic oil syndrome, eosinophilic fasciitis localized forms of scleroderma, keloid, and fibrosing colonopathy; angiomatoid malignant fibrous histiocytoma; carcinoma, including primary head and neck squamous cell carcinoma; sarcoma, including kaposi's sarcoma; fibroadanoma and phyllodes tumors, including mammary fibroadenoma; stromal tumors; phyllodes tumors, including histiocytoma; erythroblastosis; neurofibromatosis; diseases of the vascular endothelium; demyelinating, particularly in old lesions; gliosis, vasogenic edema, vascular disease, Alzheimer's and Parkinson's disease; T-cell lymphomas; B-cell lymphomas.

Disorders involving the testis and epididymis include, but are not limited to, congenital anomalies such as cryptorchidism, regressive changes such as atrophy, inflammations such as nonspecific epididymitis and orchitis, granulomatous (autoimmune) orchitis, and specific inflammations including, but not limited to, gonorrhea, mumps, tuberculosis, and syphilis, vascular disturbances including torsion,

testicular tumors including germ cell tumors that include, but are not limited to, seminoma, spermatocytic seminoma, embryonal carcinoma, yolk sac tumor choriocarcinoma, teratoma, and mixed tumors, tumore of sex cord-gonadal stroma including, but not limited to, Leydig (interstitial) cell tumors and sertoli cell tumors (androblastoma), and testicular lymphoma, and miscellaneous lesions of tunica vaginalis.

Bone-forming cells include the osteoprogenitor cells, osteoblasts, and osteocytes. The disorders of the bone are complex because they may have an impact on the skeleton during any of its stages of development. Hence, the disorders may have variable manifestations and may involve one, multiple or all bones of the body. Such disorders include, congenital malformations, achondroplasia and thanatophoric dwarfism, diseases associated with abnormal matix such as type 1 collagen disease, osteoporosis, Paget disease, rickets, osteomalacia, high-turnover osteodystrophy, low-turnover of aplastic disease, osteonecrosis, pyogenic osteomyelitis, tuberculous osteomyelitism, osteoma, osteoid osteoma, osteoblastoma, osteosarcoma, osteochondroma, chondromas, chondroblastoma, chondromyxoid fibroma, chondrosarcoma, fibrous cortical defects, fibrous dysplasia, fibrosarcoma, malignant fibrous histiocytoma, Ewing sarcoma, primitive neuroectodermal tumor, giant cell tumor, and metastatic tumors.

Disorders involving the heart, include but are not limited to, heart failure, including but not limited to, cardiac hypertrophy, left-sided heart failure, and right-sided heart failure; ischemic heart disease, including but not limited to angina pectoris, myocardial infarction, chronic ischemic heart disease, and sudden cardiac death; hypertensive heart disease, including but not limited to, systemic (left-sided) hypertensive heart disease and pulmonary (right-sided) hypertensive heart disease; valvular heart disease, including but not limited to, valvular degeneration caused by calcification, such as calcific aortic stenosis, calcification of a congenitally bicuspid aortic valve, and mitral annular calcification, and myxomatous degeneration of the mitral valve (mitral valve prolapse), rheumatic fever and rheumatic heart disease, infective endocarditis, and noninfected vegetations, such as nonbacterial thrombotic endocarditis and endocarditis of systemic lupus erythematosus (Libman-Sacks disease), carcinoid heart disease, and complications of artificial valves; myocardial disease, including but not limited to dilated cardiomyopathy, hypertrophic cardiomyopathy, restrictive cardiomyopathy, and myocarditis; pericardial disease,

including but not limited to, pericardial effusion and hemopericardium and pericarditis, including acute pericarditis and healed pericarditis, and rheumatoid heart disease; neoplastic heart disease, including but not limited to, primary cardiac tumors, such as myxoma, lipoma, papillary fibroelastoma, rhabdomyoma, and sarcoma, and cardiac effects of noncardiac neoplasms; congenital heart disease, including but not limited to, left-to-right shunts--late cyanosis, such as atrial septal defect, ventricular septal defect, patent ductus arteriosus, and atrioventricular septal defect, right-to-left shunts--early cyanosis, such as tetralogy of fallot, transposition of great arteries, truncus arteriosus, tricuspid atresia, and total anomalous pulmonary venous connection, obstructive congenital anomalies, such as coarctation of aorta, pulmonary stenosis and atresia, and aortic stenosis and atresia, and disorders involving cardiac transplantation.

Because the gene shows high expression in heart, and differential expression in normal and diseased hearts, disorders related to this tissue are particularly relevant. Cardiac hypertrophy is the principal response of the heart to overload from any cause including ischemia/reperfusion injury, myocardial infarction, longstanding heart failure, vascular wall remodeling, ventricular remodeling, dilated cardiomyopathy, rapid ventricular pacing, coronary microembolism, pressure-overload, aortic banding, coronary artery ligation, end stage heart failure, tachyarrhythmia, bradyarrhythmia, valvar heart disease, and hypertension. Hypertrophy is a strong, independent predictor of cardiovascular death and is associated with diastolic dysfunction. Since adult cardiac myocytes are terminally differentiated cells, the increase in muscle mass seen in cardiac hypertrophy occurs predominantly by an increase in myocyte size. At the cellular level, the events leading to cardiac hypertrophy have been divided into (1) extracellular hypertrophic stimulus; (2) intracellular signal transduction; and (3) activation of nuclear events that allow for the hypertrophic phenotype.

The epitope-bearing polypeptides may be produced by any conventional means (Houghten, R.A., *Proc. Natl. Acad. Sci. USA 82*:5131-5135 (1985)). Simultaneous multiple peptide synthesis is described in U.S. Patent No. 4,631,211.

Fragments can be discrete (not fused to other amino acids or polypeptides) or can be within a larger polypeptide. Further, several fragments can be comprised within a single larger polypeptide. In one embodiment a fragment designed for expression in a host can have heterologous pre- and pro-polypeptide regions fused to the amino terminus

of the polypeptide fragment and an additional region fused to the carboxyl terminus of the fragment.

The invention thus provides chimeric or fusion proteins. These comprise a 19459 protein operatively linked to a heterologous protein having an amino acid sequence not substantially homologous to the 19459 protein. "Operatively linked" indicates that the 19459 protein and the heterologous protein are fused in-frame. The heterologous protein can be fused to the N-terminus or C-terminus of the 19459 protein.

In one embodiment the fusion protein does not affect 19459 protein function *per se*. For example, the fusion protein can be a GST-fusion protein in which the 19459 sequences are fused to the N- or C-terminus of the GST sequences. Other types of fusion proteins include, but are not limited to, enzymatic fusion proteins, for example beta-galactosidase fusions, yeast two-hybrid GAL-4 fusions, poly-His fusions and Ig fusions. Such fusion proteins, particularly poly-His fusions, can facilitate the purification of recombinant 19459 protein. In certain host cells (e.g., mammalian host cells), expression and/or secretion of a protein can be increased by using a heterologous signal sequence. Therefore, in another embodiment, the fusion protein contains a heterologous signal sequence at its C- or N-terminus.

EP-A-0 464 533 discloses fusion proteins comprising various portions of immunoglobulin constant regions. The Fc is useful in therapy and diagnosis and thus results, for example, in improved pharmacokinetic properties (EP-A 0232 262). In drug discovery, for example, human proteins have been fused with Fc portions for the purpose of high-throughput screening assays to identify antagonists. Bennett *et al.* (*J. Mol. Recog. 8*:52-58 (1995)) and Johanson *et al.* (*J. Biol. Chem. 270, 16*:9459-9471 (1995)). Thus, this invention also encompasses soluble fusion proteins containing a 19459 polypeptide and various portions of the constant regions of heavy or light chains of immunoglobulins of various subclass (IgG, IgM, IgA, IgE). Preferred as immunoglobulin is the constant part of the heavy chain of human IgG, particularly IgG1, where fusion takes place at the hinge region. For some uses it is desirable to remove the Fc after the fusion protein has been used for its intended purpose, for example when the fusion protein is to be used as antigen for immunizations. In a particular embodiment, the Fc part can be removed in a simple way by a cleavage sequence which is also incorporated and can be cleaved with factor Xa.

A chimeric or fusion protein can be produced by standard recombinant DNA techniques. For example, DNA fragments coding for the different protein sequences are ligated together in-frame in accordance with conventional techniques. In another embodiment, the fusion gene can be synthesized by conventional techniques including automated DNA synthesizers. Alternatively, PCR amplification of gene fragments can be carried out using anchor primers which give rise to complementary overhangs between two consecutive gene fragments which can subsequently be annealed and re-amplified to generate a chimeric gene sequence (see Ausubel *et al.*, *Current Protocols in Molecular Biology*, 1992). Moreover, many expression vectors are commercially available that already encode a fusion moiety (e.g., a GST protein). A 19459 protein-encoding nucleic acid can be cloned into such an expression vector such that the fusion moiety is linked in-frame to the 19459 protein.

Another form of fusion protein is one that directly affects the 19459 protein functions. Accordingly, a 19459 polypeptide is encompassed by the present invention in which one or more of the domains (or parts thereof) has been replaced by homologous domains (or parts thereof) from another seven transmembrane protein, especially a G-protein coupled receptor or other type of receptor. Accordingly, various permutations are possible. The amino terminal extracellular domain, or subregion thereof, (for example, ligand-binding) can be replaced with the domain or subregion from another ligand-binding receptor protein. Alternatively, the entire transmembrane domain, or any of the seven segments or loops, or parts thereof, for example, G-protein-binding/signal transduction, can be replaced. Finally, the carboxy terminal intracellular domain or subregion can be replaced. Thus, chimeric proteins, especially receptors, can be formed in which one or more of the native domains or subregions has been replaced.

The isolated 19459 protein can be purified from cells that naturally express it, including but not limited to, those shown in Figure 5, 6, and 7, and particularly heart and kidney, purified from cells that have been altered to express it (recombinant), or synthesized using known protein synthesis methods.

In one embodiment, the protein is produced by recombinant DNA techniques. For example, a nucleic acid molecule encoding the 19459 polypeptide is cloned into an expression vector, the expression vector introduced into a host cell and the protein expressed in the host cell. The protein can then be isolated from the cells by an appropriate purification scheme using standard protein purification techniques.

Polypeptides often contain amino acids other than the 20 amino acids commonly referred to as the 20 naturally-occurring amino acids. Further, many amino acids, including the terminal amino acids, may be modified by natural processes, such as processing and other post-translational modifications, or by chemical modification techniques well known in the art. Common modifications that occur naturally in polypeptides are described in basic texts, detailed monographs, and the research literature, and they are well known to those of skill in the art.

Accordingly, the polypeptides also encompass derivatives or analogs in which a substituted amino acid residue is not one encoded by the genetic code, in which a substituent group is included, in which the mature polypeptide is fused with another compound, such as a compound to increase the half-life of the polypeptide (for example, polyethylene glycol), or in which the additional amino acids are fused to the mature polypeptide, such as a leader or secretory sequence or a sequence for purification of the mature polypeptide or a pro-protein sequence.

Known modifications include, but are not limited to, acetylation, acylation, ADP-ribosylation, amidation, covalent attachment of flavin, covalent attachment of a heme moiety, covalent attachment of a nucleotide or nucleotide derivative, covalent attachment of a lipid or lipid derivative, covalent attachment of phosphatidylinositol, cross-linking, cyclization, disulfide bond formation, demethylation, formation of covalent crosslinks, formation of cystine, formation of pyroglutamate, formylation, gamma carboxylation, glycosylation, GPI anchor formation, hydroxylation, iodination, methylation, myristoylation, oxidation, proteolytic processing, phosphorylation, prenylation, racemization, selenoylation, sulfation, transfer-RNA mediated addition of amino acids to proteins such as arginylation, and ubiquitination.

Such modifications are well-known to those of skill in the art and have been described in great detail in the scientific literature. Several particularly common modifications, glycosylation, lipid attachment, sulfation, gamma-carboxylation of glutamic acid residues, hydroxylation and ADP-ribosylation, for instance, are described in most basic texts, such as *Proteins - Structure and Molecular Properties*, 2nd Ed., T.E. Creighton, W. H. Freeman and Company, New York (1993). Many detailed reviews are available on this subject, such as by Wold, F., *Posttranslational Covalent Modification of Proteins*, B.C. Johnson, Ed., Academic Press, New York 1-12 (1983); Seifter *et al.*

(*Meth. Enzymol. 182*: 626-646 (1990)) and Rattan *et al.* (*Ann. N.Y. Acad. Sci. 663*:48-62 (1992)).

As is also well known, polypeptides are not always entirely linear. For instance, polypeptides may be branched as a result of ubiquitination, and they may be circular, with or without branching, generally as a result of post-translation events, including natural processing event and events brought about by human manipulation which do not occur naturally. Circular, branched and branched circular polypeptides may be synthesized by non-translational natural processes and by synthetic methods.

Modifications can occur anywhere in a polypeptide, including the peptide backbone, the amino acid side-chains and the amino or carboxyl termini. Blockage of the amino or carboxyl group in a polypeptide, or both, by a covalent modification, is common in naturally-occurring and synthetic polypeptides. For instance, the amino terminal residue of polypeptides made in *E. coli*, prior to proteolytic processing, almost invariably will be N-formylmethionine.

The modifications can be a function of how the protein is made. For recombinant polypeptides, for example, the modifications will be determined by the host cell posttranslational modification capacity and the modification signals in the polypeptide amino acid sequence. Accordingly, when glycosylation is desired, a polypeptide should be expressed in a glycosylating host, generally a eukaryotic cell. Insect cells often carry out the same posttranslational glycosylations as mammalian cells and, for this reason, insect cell expression systems have been developed to efficiently express mammalian proteins having native patterns of glycosylation. Similar considerations apply to other modifications.

The same type of modification may be present in the same or varying degree at several sites in a given polypeptide. Also, a given polypeptide may contain more than one type of modification.

Polypeptide uses

The 19459 polypeptides are useful for producing antibodies specific for the 19459 protein, regions, or fragments. Regions having a high antigenicity index score are shown in Figure 3.

The polypeptides (including variants and fragments which may have been disclosed prior to the present invention) are useful for biological assays related to seven

transmembrane proteins and particularly, GPCRs. Such assays involve any of the known seven transmembrane protein, and particularly GPCR, functions or activities or properties useful for diagnosis and treatment of seven transmembrane protein, and particularly GPCR-related conditions.

The 19459 polypeptides are also useful in drug screening assays, in cell-based or cell-free systems. Cell-based systems can be native, i.e., cells that normally express the protein, as a biopsy or expanded in cell culture. For the various biological assays described herein, these cells included but are not limited to, those disclosed above, and particularly heart and kidney. In one embodiment, however, cell-based assays involve recombinant host cells expressing the protein.

Determining the ability of the test compound to interact with the polypeptide can also comprise determining the ability of the test compound to preferentially bind to the polypeptide as compared to the ability of the ligand, or a biologically active portion thereof, to bind to the polypeptide.

The polypeptides can be used to identify compounds that modulate peptide, e.g., receptor, activity. Such compounds, for example, can increase or decrease affinity or rate of binding to a known ligand, compete with ligand for binding, or displace bound ligand. Both 19459 protein and appropriate variants and fragments can be used in high-throughput screens to assay candidate compounds for the ability to bind to the protein. These compounds can be further screened against a functional 19459 polypeptide to determine the effect of the compound on the 19459 protein activity. Compounds can be identified that activate (agonist) or inactivate (antagonist) the protein to a desired degree. Modulatory methods can be performed *in vitro* (e.g., by culturing the cell with the agent) or, alternatively, *in vivo* (e.g., by administering the agent to a subject).

The polypeptides can be used to screen a compound for the ability to stimulate or inhibit interaction between the protein and a target molecule that normally interacts with the protein. The target can be ligand or a component of the signal pathway with which the protein normally interacts (for example, a G-protein or other interactor involved in cAMP or phosphatidylinositol turnover and/or adenylate cyclase, or phospholipase C activation). The assay includes the steps of combining the 19459 protein with a candidate compound under conditions that allow the protein or fragment to interact with the target molecule, and to detect the formation of a complex between the protein and the target or to detect the biochemical consequence of the interaction with the protein and

the target, such as any of the associated effects of signal transduction such as G-protein phosphorylation, cyclic AMP or phosphatidylinositol turnover, and adenylate cyclase or phospholipase C activation.

Determining the ability of the protein to bind to a target molecule can also be accomplished using a technology such as real-time Bimolecular Interaction Analysis (BIA). Sjolander, S. and Urbaniczky, C. (1991) *Anal. Chem. 63*:2338-2345 and Szabo *et al.* (1995) *Curr. Opin. Struct. Biol. 5*:699-705. As used herein, "BIA" is a technology for studying biospecific interactions in real time, without labeling any of the interactants (e.g., BIAcore™). Changes in the optical phenomenon surface plasmon resonance (SPR) can be used as an indication of real-time reactions between biological molecules.

The test compounds of the present invention can be obtained using any of the numerous approaches in combinatorial library methods known in the art, including: biological libraries; spatially addressable parallel solid phase or solution phase libraries; synthetic library methods requiring deconvolution; the 'one-bead one-compound' library method; and synthetic library methods using affinity chromatography selection. The biological library approach is limited to polypeptide libraries, while the other four approaches are applicable to polypeptide, non-peptide oligomer or small molecule libraries of compounds (Lam, K.S. (1997) *Anticancer Drug Des. 12*:145).

Examples of methods for the synthesis of molecular libraries can be found in the art, for example in DeWitt *et al.* (1993) *Proc. Natl. Acad. Sci. USA 90*:6909; Erb *et al.* (1994) *Proc. Natl. Acad. Sci. USA 91*:11422; Zuckermann *et al.* (1994). *J. Med. Chem. 37*:2678; Cho *et al.* (1993) *Science 261*:1303; Carell *et al.* (1994) *Angew. Chem. Int. Ed. Engl. 33*:2059; Carell *et al.* (1994) *Angew. Chem. Int. Ed. Engl. 33*:2061; and in Gallop *et al.* (1994) *J. Med. Chem. 37*:1233. Libraries of compounds may be presented in solution (e.g., Houghten (1992) *Biotechniques 13*:412-421), or on beads (Lam (1991) *Nature 354*:82-84), chips (Fodor (1993) *Nature 364*:555-556), bacteria (Ladner USP 5,223,409), spores (Ladner USP '409), plasmids (Cull *et al.* (1992) *Proc. Natl. Acad. Sci. USA 89*:1865-1869) or on phage (Scott and Smith (1990) *Science 249*:386-390); (Devlin (1990) *Science 249*:404-406); (Cwirla *et al.* (1990) *Proc. Natl. Acad. Sci. 97*:6378-6382); (Felici (1991) *J. Mol. Biol. 222*:301-310); (Ladner *supra*).

Candidate compounds include, for example, 1) peptides such as soluble peptides, including Ig-tailed fusion peptides and members of random peptide libraries (see, e.g., Lam *et al.*, *Nature 354*:82-84 (1991); Houghten *et al.*, *Nature 354*:84-86 (1991)) and combinatorial chemistry-derived molecular libraries made of D- and/or L- configuration amino acids; 2) phosphopeptides (e.g., members of random and partially degenerate, directed phosphopeptide libraries, see, e.g., Songyang *et al.*, *Cell 72*:767-778 (1993)); 3) antibodies (e.g., polyclonal, monoclonal, humanized, anti-idiotypic, chimeric, and single chain antibodies as well as Fab, $F(ab')_2$, Fab expression library fragments, and epitope-binding fragments of antibodies); and 4) small organic and inorganic molecules (e.g., molecules obtained from combinatorial and natural product libraries).

One candidate compound is a soluble full-length 19459 protein or fragment that competes for ligand binding. Other candidate compounds include mutant 19459 proteins or appropriate fragments containing mutations that affect protein function and thus compete for ligand. Accordingly, a fragment that competes for ligand, for example with a higher affinity, or a fragment that binds ligand but does not allow release, is encompassed by the invention.

The invention provides other end points to identify compounds that modulate (stimulate or inhibit) protein activity. The assays typically involve an assay of events in the signal transduction pathway that indicate receptor activity. Thus, the expression of genes that are up- or down-regulated in response to the receptor protein dependent signal cascade can be assayed. In one embodiment, the regulatory region of such genes can be operably linked to a marker that is easily detectable, such as luciferase. Alternatively, phosphorylation of the 19459 protein, or a 19459 protein target, could also be measured.

Any of the biological or biochemical functions mediated by the 19459 protein can be used as an endpoint assay. These include all of the biochemical or biochemical/biological events described herein, in the references cited herein, incorporated by reference for these endpoint assay targets, and other functions known to those of ordinary skill in the art.

Binding and/or activating compounds can also be screened by using chimeric 19459 proteins in which the amino terminal extracellular domain, or parts thereof, the entire transmembrane domain or subregions, such as any of the seven transmembrane segments or any of the intracellular or extracellular loops and the carboxy terminal intracellular domain, or parts thereof, can be replaced by heterologous domains or

subregions. For example, a G-protein-binding region can be used that interacts with a different G-protein than that which is recognized by the native receptor. Accordingly, a different set of signal transduction components is available as an end-point assay for activation. Alternatively, the entire transmembrane portion or subregions (such as transmembrane segments or intracellular or extracellular loops) can be replaced with the entire transmembrane portion or subregions specific to a host cell that is different from the host cell from which the amino terminal extracellular domain and/or the G-protein-binding region are derived. This allows for assays to be performed in other than the specific host cell from which the 19459 protein is derived. Alternatively, the amino terminal extracellular domain (and/or other ligand-binding regions) could be replaced by a domain (and/or other binding region) binding a different ligand, thus, providing an assay for test compounds that interact with the heterologous amino terminal extracellular domain (or region) but still cause signal transduction. Finally, activation can be detected by a reporter gene containing an easily detectable coding region operably linked to a transcriptional regulatory sequence that is part of the native signal transduction pathway.

The 19459 polypeptides are also useful in competition binding assays in methods designed to discover compounds that interact with the polypeptide. Thus, a compound is exposed to the polypeptide under conditions that allow the compound to bind or to otherwise interact with the polypeptide. Soluble 19459 polypeptide is also added to the mixture. If the test compound interacts with the soluble polypeptide, it decreases the amount of complex formed or activity from the target. This type of assay is particularly useful in cases in which compounds are sought that interact with specific regions of the polypeptide. Thus, the soluble polypeptide that competes with the target region is designed to contain peptide sequences corresponding to the region of interest.

To perform cell free drug screening assays, it is desirable to immobilize either the protein, or fragment, or its target molecule to facilitate separation of complexes from uncomplexed forms of one or both of the proteins, as well as to accommodate automation of the assay.

Techniques for immobilizing proteins on matrices can be used in the drug screening assays. In one embodiment, a fusion protein can be provided which adds a domain that allows the protein to be bound to a matrix. For example, glutathione-S-transferase/19459 fusion proteins can be adsorbed onto glutathione sepharose beads (Sigma Chemical, St. Louis, MO) or glutathione derivatized microtitre plates, which are

then combined with the cell lysates (e.g., ^{35}S-labeled) and the candidate compound, and the mixture incubated under conditions conducive to complex formation (e.g., at physiological conditions for salt and pH). Following incubation, the beads are washed to remove any unbound label, and the matrix immobilized and radiolabel determined directly, or in the supernatant after the complexes are dissociated. Alternatively, the complexes can be dissociated from the matrix, separated by SDS-PAGE, and the level of 19459 peptide-binding protein found in the bead fraction quantitated from the gel using standard electrophoretic techniques. For example, either the polypeptide or its target molecule can be immobilized utilizing conjugation of biotin and streptavidin using techniques well known in the art. Alternatively, antibodies reactive with the protein but which do not interfere with binding of the protein to its target molecule can be derivatized to the wells of the plate, and the protein trapped in the wells by antibody conjugation. Preparations of a 19459 peptide-binding protein and a candidate compound are incubated in the 19459 protein-presenting wells and the amount of complex trapped in the well can be quantitated. Methods for detecting such complexes, in addition to those described above for the GST-immobilized complexes, include immunodetection of complexes using antibodies reactive with the 19459 protein target molecule, or which are reactive with 19459 protein and compete with the target molecule; as well as enzyme-linked assays which rely on detecting an enzymatic activity associated with the target molecule.

Modulators of 19459 protein activity identified according to these drug screening assays can be used to treat a subject with a disorder mediated by the 19459 protein pathway, by treating cells that express the 19459 protein, such as those disclosed herein.

Preferred disorders include, but are not limited to, those involving the heart, such as are mentioned above.

These methods of treatment include the steps of administering the modulators of protein activity in a pharmaceutical composition as described herein, to a subject in need of such treatment.

The 19459 polypeptides are thus useful for treating a 19459 protein-associated disorder characterized by aberrant expression or activity of a 19459 protein. In one embodiment, the method involves administering an agent (e.g., an agent identified by a screening assay described herein), or combination of agents that modulates (e.g., upregulates or downregulates) expression or activity of the protein. In another

embodiment, the method involves administering a protein as therapy to compensate for reduced or aberrant expression or activity of the protein.

Stimulation of protein activity is desirable in situations in which the protein is abnormally downregulated and/or in which increased protein activity is likely to have a beneficial effect. Likewise, inhibition of protein activity is desirable in situations in which the protein is abnormally upregulated and/or in which decreased protein activity is likely to have a beneficial effect. In one example of such a situation, a subject has a disorder characterized by aberrant development or cellular differentiation. In another example of such a situation, the subject has a proliferative disease (e.g., cancer) or a disorder characterized by an aberrant hematopoietic response. In another example of such a situation, it is desirable to achieve tissue regeneration in a subject (e.g., where a subject has undergone brain or spinal cord injury and it is desirable to regenerate neuronal tissue in a regulated manner).

In yet another aspect of the invention, the proteins of the invention can be used as "bait proteins" in a two-hybrid assay or three-hybrid assay (see, e.g., U.S. Patent No. 5,283,317; Zervos *et al.* (1993) *Cell 72*:223-232; Madura *et al.* (1993) *J. Biol. Chem. 268*:12046-12054; Bartel *et al.* (1993) *Biotechniques 14*:920-924; Iwabuchi *et al.* (1993) *Oncogene 8*:1693-1696; and Brent WO 94/10300), to identify other proteins (captured proteins) which bind to or interact with the proteins of the invention and modulate their activity.

The 19459 polypeptides also are useful to provide a target for diagnosing a disease or predisposition to disease mediated by the 19459 protein, especially in diseases involving the tissues in which the 19459 protein is expressed as disclosed herein. Accordingly, methods are provided for detecting the presence, or levels of, the 19459 protein in a cell, tissue, or organism. The method involves contacting a biological sample with a compound capable of interacting with the protein such that the interaction can be detected.

One agent for detecting the protein is an antibody capable of selectively binding to the protein. A biological sample includes tissues, cells and biological fluids isolated from a subject, as well as tissues, cells and fluids present within a subject.

The 19459 protein also provides a target for diagnosing active disease, or predisposition to disease, in a patient having a variant 19459 protein. Thus, the protein can be isolated from a biological sample, assayed for the presence of a genetic mutation

that results in an aberrant protein. This includes amino acid substitution, deletion, insertion, rearrangement, (as the result of aberrant splicing events), and inappropriate post-translational modification. Analytic methods include altered electrophoretic mobility, altered tryptic peptide digest, altered 19459 activity in cell-based or cell-free assays, alteration in ligand or antibody-binding pattern, altered isoelectric point, direct amino acid sequencing, and any other of the known assay techniques useful for detecting mutations in a protein.

In vitro techniques for detection of 19459 protein include enzyme linked immunosorbent assays (ELISAs), Western blots, immunoprecipitations and immunofluorescence. Alternatively, the protein can be detected *in vivo* in a subject by introducing into the subject a labeled anti-19459 protein antibody. For example, the antibody can be labeled with a radioactive marker whose presence and location in a subject can be detected by standard imaging techniques. Particularly useful are methods which detect the allelic variant of the protein expressed in a subject and methods which detect fragments of the protein in a sample.

The polypeptides are also useful in pharmacogenomic analysis. Pharmacogenomics deal with clinically significant hereditary variations in the response to drugs due to altered drug disposition and abnormal action in affected persons. See, e.g., Eichelbaum, M., *Clin. Exp. Pharmacol. Physiol.* 23(10-11) :983-985 (1996), and Linder, M.W., *Clin. Chem.* 43(2):254-266 (1997). The clinical outcomes of these variations result in severe toxicity of therapeutic drugs in certain individuals or therapeutic failure of drugs in certain individuals as a result of individual variation in metabolism. Thus, the genotype of the individual can determine the way a therapeutic compound acts on the body or the way the body metabolizes the compound. Further, the activity of drug metabolizing enzymes effects both the intensity and duration of drug action. Thus, the pharmacogenomics of the individual permit the selection of effective compounds and effective dosages of such compounds for prophylactic or therapeutic treatment based on the individual's genotype. The discovery of genetic polymorphisms in some drug metabolizing enzymes has explained why some patients do not obtain the expected drug effects, show an exaggerated drug effect, or experience serious toxicity from standard drug dosages. Polymorphisms can be expressed in the phenotype of the extensive metabolizer and the phenotype of the poor metabolizer. Accordingly, genetic polymorphism may lead to allelic protein variants in which one or more functions in one

population is different from those in another population. The polypeptides thus allow a target to ascertain a genetic predisposition that can affect treatment modality. Thus, in a ligand-based treatment, polymorphism may give rise to amino terminal extracellular domains and/or other ligand-binding regions that are more or less active in ligand binding and receptor activation. Accordingly, ligand dosage would necessarily be modified to maximize the therapeutic effect within a given population containing a polymorphism. As an alternative to genotyping, specific polymorphic polypeptides could be identified.

The polypeptides are also useful for monitoring therapeutic effects during clinical trials and other treatment. Thus, the therapeutic effectiveness of an agent that is designed to increase or decrease gene expression, protein levels or activity can be monitored over the course of treatment using the polypeptides as an end-point target. The monitoring can be, for example, as follows: (i) obtaining a pre-administration sample from a subject prior to administration of the agent; (ii) detecting the level of expression or activity of a specified protein in the pre-administration sample; (iii) obtaining one or more post-administration samples from the subject; (iv) detecting the level of expression or activity of the protein in the post-administration samples; (v) comparing the level of expression or activity of the protein in the pre-administration sample with the protein in the post-administration sample or samples; and (vi) increasing or decreasing the administration of the agent to the subject accordingly.

The polypeptides are also useful for treating a 19459 protein-associated disorder. Accordingly, methods for treatment include the use of soluble protein or fragments of the protein that compete for ligand binding. These proteins or fragments can have a higher affinity for the ligand so as to provide effective competition.

Antibodies

The invention also provides antibodies that selectively bind to the 19459 protein and its variants and fragments. An antibody is considered to selectively bind, even if it also binds to other proteins that are not substantially homologous with the 19459 protein. These other proteins share homology with a fragment or domain of the protein. This conservation in specific regions gives rise to antibodies that bind to both proteins by virtue of the homologous sequence. In this case, it would be understood that antibody binding to the protein is still selective.

To generate antibodies, an isolated polypeptide is used as an immunogen to generate antibodies using standard techniques for polyclonal and monoclonal antibody preparation. Either the full-length protein or antigenic peptide fragment can be used. Regions having a high antigenicity index are shown in Figure 3.

Antibodies are preferably prepared from these regions or from discrete fragments in these regions. However, antibodies can be prepared from any region of the peptide as described herein. A preferred fragment produces an antibody that diminishes or completely prevents ligand-binding. Antibodies can be developed against the entire 19459 protein or portions of the protein, for example, the intracellular carboxy terminal domain, the amino terminal extracellular domain, the entire transmembrane domain or specific segments, any of the intra or extracellular loops, or any portions of the above. Antibodies may also be developed against specific functional sites, such as the site of ligand-binding, the site of G protein coupling, or sites that are phosphorylated, myristoylated, or glycosylated.

An antigenic fragment will typically comprise at least 6 contiguous amino acid residues. The antigenic peptide can comprise a contiguous sequence of at least 12, at least 14 amino acid residues, at least 15 amino acid residues, at least 20 amino acid residues, or at least 30 amino acid residues. In one embodiment, fragments correspond to regions that are located on the surface of the protein, e.g., hydrophilic regions. These fragments are not to be construed, however, as encompassing any fragments which may be disclosed prior to the invention.

Antibodies can be polyclonal or monoclonal. An intact antibody, or a fragment thereof (e.g. Fab or $F(ab')_2$) can be used.

Detection can be facilitated by coupling (i.e., physically linking) the antibody to a detectable substance. Examples of detectable substances include various enzymes, prosthetic groups, fluorescent materials, luminescent materials, bioluminescent materials, and radioactive materials. Examples of suitable enzymes include horseradish peroxidase, alkaline phosphatase, β-galactosidase, or acetylcholinesterase; examples of suitable prosthetic group complexes include streptavidin/biotin and avidin/biotin; examples of suitable fluorescent materials include umbelliferone, fluorescein, fluorescein isothiocyanate, rhodamine, dichlorotriazinylamine fluorescein, dansyl chloride or phycoerythrin; an example of a luminescent material includes luminol;

examples of bioluminescent materials include luciferase, luciferin, and aequorin, and examples of suitable radioactive material include ^{125}I, ^{131}I, ^{35}S or ^{3}H.

An appropriate immunogenic preparation can be derived from native, recombinantly expressed, protein or chemically synthesized peptides.

Antibody Uses

The antibodies can be used to isolate a protein by standard techniques, such as affinity chromatography or immunoprecipitation. The antibodies can facilitate the purification of the natural protein from cells and recombinantly produced protein expressed in host cells.

The antibodies are useful to detect the presence of the protein in cells or tissues to determine the pattern of expression among various tissues in an organism and over the course of normal development.

The antibodies can be used to detect the protein *in situ*, *in vitro*, or in a cell lysate or supernatant in order to evaluate the abundance and pattern of expression.

The antibodies can be used to assess abnormal tissue distribution or abnormal expression during development.

Antibody detection of circulating fragments of the full length 19459 protein can be used to identify protein turnover.

Further, the antibodies can be used to assess 19459 protein expression in disease states such as in active stages of the disease or in an individual with a predisposition toward disease related to 19459 protein function. When a disorder is caused by an inappropriate tissue distribution, developmental expression, or level of expression of the protein, the antibody can be prepared against the normal protein. If a disorder is characterized by a specific mutation in the protein, antibodies specific for this mutant protein can be used to assay for the presence of the specific mutant protein. However, intracellularly-made antibodies ("intrabodies") are also encompassed, which would recognize intracellular peptide regions.

The antibodies can also be used to assess normal and aberrant subcellular localization of cells in the various tissues in an organism. Antibodies can be developed against the whole protein or portions, for example, portions of the amino terminal extracellular domain or extracellular loops.

The diagnostic uses can be applied, not only in genetic testing, but also in monitoring a treatment modality. Accordingly, where treatment is ultimately aimed at correcting 19459 protein expression level or the presence of aberrant 19459 protein and aberrant tissue distribution or developmental expression, antibodies directed against the protein or relevant fragments can be used to monitor therapeutic efficacy. Antibodies accordingly can be used diagnostically to monitor protein levels in tissue as part of a clinical testing procedure, e.g., to, for example, determine the efficacy of a given treatment regimen.

Additionally, antibodies are useful in pharmacogenomic analysis. Thus, antibodies prepared against polymorphic 19459 proteins can be used to identify individuals that require modified treatment modalities.

The antibodies are also useful as diagnostic tools as an immunological marker for aberrant protein analyzed by electrophoretic mobility, isoelectric point, tryptic peptide digest, and other physical assays known to those in the art.

The antibodies are also useful for tissue typing. Thus, where a specific 19459 protein has been correlated with expression in a specific tissue, antibodies that are specific for this protein can be used to identify a tissue type.

The antibodies are also useful in forensic identification. Accordingly, where an individual has been correlated with a specific genetic polymorphism resulting in a specific polymorphic protein, an antibody specific for the polymorphic protein can be used as an aid in identification.

The antibodies are also useful for inhibiting protein function, for example, blocking ligand binding.

These uses can also be applied in a therapeutic context in which treatment involves inhibiting a function. An antibody can be used, for example, to block ligand binding. Antibodies can be prepared against specific fragments containing sites required for function or against the intact 19459 protein associated with a cell.

Completely human antibodies are particularly desirable for therapeutic treatment of human patients. For an overview of this technology for producing human antibodies, see Lonberg and Huszar (1995, *Int. Rev. Immunol. 13*:65-93). For a detailed discussion of this technology for producing human antibodies and human monoclonal antibodies and protocols for producing such antibodies, see, e.g., U.S.

Patent 5,625,126; U.S. Patent 5,633,425; U.S. Patent 5,569,825; U.S. Patent 5,661,016; and U.S. Patent 5,545,806.

The invention also encompasses kits for using antibodies to detect the presence of a 19459 protein in a biological sample. The kit can comprise antibodies such as a labeled or labelable antibody and a compound or agent for detecting the protein in a biological sample; means for determining the amount of the protein in the sample; and means for comparing the amount of the protein in the sample with a standard. The compound or agent can be packaged in a suitable container. The kit can further comprise instructions for using the kit to detect the protein.

Polynucleotides

The nucleotide sequence in SEQ ID NO:1 was obtained by sequencing the deposited human full length cDNA. Accordingly, the sequence of the deposited clone is controlling as to any discrepancies between the two and any reference to the sequence of SEQ ID NO:1 includes reference to the sequence of the deposited cDNA.

The specifically disclosed cDNA comprises the coding region and 5' and 3' untranslated sequences (SEQ ID NO:1).

The human 19459 cDNA is approximately 1737 nucleotides in length and encodes a full length protein that is approximately 397 amino acid residues in length. The nucleic acid is expressed in tissues including, but not limited to, those described herein above. Structural analysis of the amino acid sequence of SEQ ID NO:2 is provided in Figure 2, a hydropathy plot. The figure shows the putative structure of the seven transmembrane segments, the amino terminal extracellular domain and the carboxy terminal intracellular domain.

As used herein, the term "transmembrane segment" refers to a structural amino acid motif which includes a hydrophobic helix that spans the plasma membrane.

The invention provides isolated polynucleotides encoding a 19459 protein. The term "19459 polynucleotide" or "19459 nucleic acid" refers to the sequence shown in SEQ ID NO:1, SEQ ID NO:3, or in the deposited cDNA. The terms further include variants and fragments of the 19459 polynucleotide.

An "isolated" nucleic acid is one that is separated from other nucleic acid present in the natural source of the nucleic acid. Preferably, an "isolated" nucleic acid is free of sequences which naturally flank the nucleic acid (i.e., sequences located at the 5' and 3'

ends of the nucleic acid) in the genomic DNA of the organism from which the nucleic acid is derived. However, there can be some flanking nucleotide sequences, for example up to about 5KB. The important point is that the nucleic acid is isolated from flanking sequences such that it can be subjected to the specific manipulations described herein such as recombinant expression, preparation of probes and primers, and other uses specific to the 19459 nucleic acid sequences.

Moreover, an "isolated" nucleic acid molecule, such as a cDNA or RNA molecule, can be substantially free of other cellular material, or culture medium when produced by recombinant techniques, or chemical precursors or other chemicals when chemically synthesized. However, the nucleic acid molecule can be fused to other coding or regulatory sequences and still be considered isolated.

For example, recombinant DNA molecules contained in a vector are considered isolated. Further examples of isolated DNA molecules include recombinant DNA molecules maintained in heterologous host cells or purified (partially or substantially) DNA molecules in solution. Isolated RNA molecules include *in vivo* or *in vitro* RNA transcripts of the isolated DNA molecules of the present invention. Isolated nucleic acid molecules according to the present invention further include such molecules produced synthetically.

In some instances, the isolated material will form part of a composition (for example, a crude extract containing other substances), buffer system or reagent mix. In other circumstances, the material may be purified to essential homogeneity, for example as determined by PAGE or column chromatography such as HPLC. Preferably, an isolated nucleic acid comprises at least about 50, 80 or 90 % (on a molar basis) of all macromolecular species present.

The 19459 polynucleotides can encode the mature protein plus additional amino or carboxyl-terminal amino acids, or amino acids interior to the mature polypeptide (when the mature form has more than one polypeptide chain, for instance). Such sequences may play a role in processing of a protein from precursor to a mature form, facilitate protein trafficking, prolong or shorten protein half-life or facilitate manipulation of a protein for assay or production, among other things. As generally is the case *in situ*, the additional amino acids may be processed away from the mature protein by cellular enzymes.

The 19459 polynucleotides include, but are not limited to, the sequence encoding the mature polypeptide alone, the sequence encoding the mature polypeptide and additional coding sequences, such as a leader or secretory sequence (e.g., a pre-pro or pro-protein sequence), the sequence encoding the mature polypeptide, with or without the additional coding sequences, plus additional non-coding sequences, for example introns and non-coding 5' and 3' sequences such as transcribed but non-translated sequences that play a role in transcription, mRNA processing (including splicing and polyadenylation signals), ribosome binding and stability of mRNA. In addition, the polynucleotide may be fused to a marker sequence encoding, for example, a peptide that facilitates purification.

Polynucleotides can be in the form of RNA, such as mRNA, or in the form of DNA, including cDNA and genomic DNA obtained by cloning or produced by chemical synthetic techniques or by a combination thereof. The nucleic acid, especially DNA, can be double-stranded or single-stranded. Single-stranded nucleic acid can be the coding strand (sense strand) or the non-coding strand (anti-sense strand).

One nucleic acid comprises the full length nucleotide sequence shown in SEQ ID NO:1, corresponding to human cDNA.

In one embodiment, the nucleic acid comprises only the coding region (SEQ ID NO:3).

The invention further provides variant polynucleotides, and fragments thereof, that differ from the nucleotide sequence shown in SEQ ID NO:1 and SEQ ID NO:3 due to degeneracy of the genetic code and thus encode the same protein as that encoded by the nucleotide sequence shown in SEQ ID NO:1 or SEQ ID NO:3.

The invention also provides nucleic acid molecules encoding the variant polypeptides described herein. Such polynucleotides may be naturally occurring, such as allelic variants (same locus), homologs (different locus), and orthologs (different organism), or may be constructed by recombinant DNA methods or by chemical synthesis. Such non-naturally occurring variants may be made by mutagenesis techniques, including those applied to polynucleotides, cells, or organisms. Accordingly, as discussed above, the variants can contain nucleotide substitutions, deletions, inversions and insertions.

Variation can occur in either or both the coding and non-coding regions. The variations can produce both conservative and non-conservative amino acid substitutions.

Typically, variants have a substantial identity with the nucleic acid molecule of SEQ ID NO:1, SEQ ID NO:3, or complements thereof.

Orthologs, homologs, and allelic variants can be identified using methods well known in the art. These variants comprise a nucleotide sequence encoding a protein that is at least about 55%, 60%, 65%, 70%, typically at least about 75%, more typically at least about 80%, 85%, 90%, and most typically at least about 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99% or more homologous to the nucleotide sequence shown in SEQ ID NO:1, SEQ ID NO:3 or a fragment of one of these sequence. Such nucleic acid molecules can readily be identified as being able to hybridize under stringent conditions, to the nucleotide sequence shown in SEQ ID NO:1, SEQ ID NO:3, or a fragment of one of these sequence. It is understood that stringent hybridization does not indicate substantial homology where it is due to general homology, such as poly A sequences, or sequences common to all or most proteins, all or most seven transmembrane proteins, or all GPCRs. Moreover, it is understood that variants do not include any of the nucleic acid sequences that may have been disclosed prior to the invention.

As used herein, the term "hybridizes under stringent conditions" describes conditions for hybridization and washing. Stringent conditions are known to those skilled in the art and can be found in Current Protocols in Molecular Biology John Wiley & Sons, N.Y. (1989), 6.3.1-6.3.6. Aqueous and nonaqueous methods are described in that reference and either can be used. A preferred, example of stringent hybridization conditions are hybridization in 6X sodium chloride/sodium citrate (SSC) at about 45°C, followed by one or more washes in 0.2X SSC, 0.1% SDS at 50°C. Another example of stringent hybridization conditions are hybridization in 6X sodium chloride/sodium citrate (SSC) at about 45°C, followed by one or more washes in 0.2X SSC, 0.1% SDS at 55°C. A further example of stringent hybridization conditions are hybridization in 6X sodium chloride/sodium citrate (SSC) at about 45°C, followed by one or more washes in 0.2X SSC, 0.1% SDS at 60°C. Preferably, stringent hybridization conditions are hybridization in 6X sodium chloride/sodium citrate (SSC) at about 45°C, followed by one or more washes in 0.2X SSC, 0.1% SDS at 65°C. Particularly preferred stringency conditions (and the conditions that should be used if the practitioner is uncertain about what conditions should be applied to determine if a molecule is within a hybridization limitation of the invention) are 0.5M Sodium Phosphate, 7% SDS at 65°C, followed by one or more washes at 0.2X SSC, 1% SDS at 65°C. Preferably, an isolated nucleic acid

molecule of the invention that hybridizes under stringent conditions to the sequence of SEQ ID NO:1 or SEQ ID NO:3, corresponds to a naturally-occurring nucleic acid molecule.

As used herein, a "naturally-occurring" nucleic acid molecule refers to an RNA or DNA molecule having a nucleotide sequence that occurs in nature (e.g., encodes a natural protein).

As understood by those of ordinary skill, the exact conditions can be determined empirically and depend on ionic strength, temperature and the concentration of destabilizing agents such as formamide or denaturing agents such as SDS. Other factors considered in determining the desired hybridization conditions include the length of the nucleic acid sequences, base composition, percent mismatch between the hybridizing sequences and the frequency of occurrence of subsets of the sequences within other non-identical sequences. Thus, equivalent conditions can be determined by varying one or more of these parameters while maintaining a similar degree of identity or similarity between the two nucleic acid molecules.

The present invention also provides isolated nucleic acids that contain a single or double stranded fragment or portion that hybridizes under stringent conditions to the nucleotide sequence of SEQ ID NO:1, SEQ ID NO:3, and the complements of thereof. In one embodiment, the nucleic acid consists of a portion of the nucleotide sequence of SEQ ID NO:1, SEQ ID NO:3, and the complements thereof. The nucleic acid fragments of the invention are at least about 10, 15, preferably at least about 20 or 25 nucleotides, and can be 30, 40, 50, 75, 100, 150, 200, 250, 300, 400, 500 or more nucleotides in length. Longer fragments, for example, 30 or more nucleotides in length, which encode antigenic proteins or polypeptides described herein are useful.

Furthermore, the invention provides polynucleotides that comprise a fragment of the full length 19459 polynucleotides. The fragment can be single or double stranded and can comprise DNA or RNA. The fragment can be derived from either the coding or the non-coding sequence.

Alternatively, a nucleic acid molecule that is a fragment of a 19450 nucleotide sequence of the present invention comprises a nucleotide sequence consisting of nucleotides 1-508, 1161-1305, 1491-1737, 1-100, 101-200, 201-300, 301-400, 401-500, 501-600, 601-700, 701-800, 801-900, 901-1000, 1001-1100, 1101-1200, 1201-1300, 1301-1400, 1401-1500, 1501-1600, 1601-1700, or 1701-1737 of SEQ ID NO:1.

In another embodiment an isolated nucleic acid encodes the entire coding region from amino acid 1 to amino acid 397. In another embodiment the isolated nucleic acid encodes a sequence corresponding to the mature protein from about amino acid 44 to amino acid 397. Other fragments include nucleotide sequences encoding the amino acid fragments described herein. Further fragments can include subfragments of the specific domains or sites described herein. Fragments also include nucleic acid sequences corresponding to specific amino acid sequences described above or fragments thereof. Nucleic acid fragments, according to the present invention, are not to be construed as encompassing those fragments that may have been disclosed prior to the invention.

Nucleic acid fragments further include sequences corresponding to the domains described herein, subregions also described, and specific functional sites. Nucleic acid fragments also include combinations of the domains, segments, loops, and other functional sites described above. Thus, for example, a 19459 nucleic acid could include sequences corresponding to the amino terminal extracellular domain and one transmembrane fragment. A person of ordinary skill in the art would be aware of the many permutations that are possible.

Where the location of the domains or sites have been predicted by computer analysis, one of ordinary skill would appreciate that the amino acid residues constituting these domains can vary depending on the criteria used to define the domains.

However, it is understood that a fragment includes any nucleic acid sequence that does not include the entire gene.

Nucleic acid fragments thus include nucleic acid molecules encoding a polypeptide comprising the amino terminal extracellular domain, a polypeptide comprising the region spanning the transmembrane domain, a polypeptide comprising the carboxy terminal intracellular domain, nucleic acid molecules encoding any of the seven transmembrane segments, extracellular or intracellular loops, or any of the functional sites disclosed herein, such as glycosylation, phosphorylation, or myristoylation sites. Where the location of the domains have been predicted by computer analysis, one of ordinary skill would appreciate that the amino acid residues constituting these domains can vary depending on the criteria used to define the domains.

The invention also provides nucleic acid fragments that encode epitope bearing regions of the proteins described herein.

The isolated polynucleotide sequences, and especially fragments, are useful as DNA probes and primers.

For example, the coding region of a 19459 gene can be isolated using the known nucleotide sequence to synthesize an oligonucleotide probe. A labeled probe can then be used to screen a cDNA library, genomic DNA library, or mRNA to isolate nucleic acid corresponding to the coding region. Further, primers can be used in PCR reactions to clone specific regions of 19459 genes.

A probe/primer typically comprises substantially purified oligonucleotide. The oligonucleotide typically comprises a region of nucleotide sequence that hybridizes under stringent conditions to at least about 5, 10, 12, typically about 25, more typically about 40, 50 or 75 consecutive nucleotides of SEQ ID NO:1 or SEQ ID NO:3 sense or anti-sense strand or other 19459 polynucleotides. A probe further comprises a label, e.g., radioisotope, fluorescent compound, enzyme, or enzyme co-factor.

Polynucleotide Uses

The nucleic acid sequences of the present invention can be used as a "query sequence" to perform a search against public databases to, for example, identify other family members or related sequences. Such searches can be performed using the NBLAST and XBLAST programs (version 2.0) of Altschul *et al.* (1990) *J. Mol. Biol. 215*:403-10. BLAST nucleotide searches can be performed with the NBLAST program, score = 100, wordlength = 12 to obtain nucleotide sequences homologous to the nucleic acid molecules of the invention. To obtain gapped alignments for comparison purposes, Gapped BLAST can be utilized as described in Altschul *et al.* (1997) *Nucleic Acids Res. 25*(17):3389-3402. When utilizing BLAST and Gapped BLAST programs, the default parameters of the respective programs (e.g., XBLAST and NBLAST) can be used. See http://www.ncbi.nlm.nih.gov.

The nucleic acid fragments of the invention provide probes or primers in assays such as those described below. "Probes" are oligonucleotides that hybridize in a base-specific manner to a complementary strand of nucleic acid. Such probes include polypeptide nucleic acids, as described in Nielsen *et al.* (1991) *Science 254*:1497-1500. Typically, a probe comprises a region of nucleotide sequence that hybridizes under highly stringent conditions to at least about 15, typically about 20-25, and more typically about 40, 50 or 75 consecutive nucleotides of the nucleic acid

of SEQ ID NO:1, SEQ ID NO:3, or the complements thereof. More typically, the probe further comprises a label, e.g., radioisotope, fluorescent compound, enzyme, or enzyme co-factor.

As used herein, the term "primer" refers to a single-stranded oligonucleotide which acts as a point of initiation of template-directed DNA synthesis using well-known methods (e.g., PCR, LCR) including, but not limited to those described herein. The appropriate length of the primer depends on the particular use, but typically ranges from about 15 to 30 nucleotides. The term "primer site" refers to the area of the target DNA to which a primer hybridizes. The term "primer pair" refers to a set of primers including a 5' (upstream) primer that hybridizes with the 5' end of the nucleic acid sequence to be amplified and a 3' (downstream) primer that hybridizes with the complement of the sequence to be amplified.

The polynucleotides are useful for probes, primers, and in biological assays, including, but not limited to, methods using the cells and tissues in which the gene is expressed, particularly in which the gene is significantly expressed, and involving disorders including, but not limited to, those also discussed herein above with respect to biological methods and assays involving the 19459 polypeptides.

Where the polynucleotides are used to assess seven transmembrane protein properties, and especially GPCR properties or functions, such as in the assays described herein, all or less than all of the entire cDNA can be useful. In this case, even fragments that may have been known prior to the invention are encompassed. Thus, for example, assays specifically directed to seven transmembrane proteins, and especially GPCR functions, such as assessing agonist or antagonist activity, encompass the use of known fragments. Further, diagnostic methods for assessing function can also be practiced with any fragment, including those fragments that may have been known prior to the invention. Similarly, in methods involving modulation or treatment of 19459-related dysfunction, all fragments are encompassed including those which may have been known in the art.

The polynucleotides are useful as a hybridization probe for cDNA and genomic DNA to isolate a full-length cDNA and genomic clones encoding the polypeptide described in SEQ ID NO:2 and to isolate cDNA and genomic clones that correspond to variants producing the same polypeptide shown in SEQ ID NO:2 or the other variants described herein. Variants can be isolated from the same tissue and organism from

which the polypeptide shown in SEQ ID NO:2 was isolated, different tissues from the same organism, or from different organisms. This method is useful for isolating genes and cDNA that are developmentally-controlled and therefore may be expressed in the same tissue or different tissues at different points in the development of an organism.

The probe can correspond to any sequence along the entire length of the gene encoding the 19459 protein. Accordingly, it could be derived from 5' noncoding regions, the coding region, and 3' noncoding regions. It is understood, however, as discussed herein, that fragments corresponding to the probe do not include those fragments that may have been disclosed prior to the present invention.

The nucleic acid probe can be, for example, the full-length cDNA of SEQ ID NO:1, or a fragment thereof, such as an oligonucleotide of at least 5, 10, 12, 15, 30, 50, 100, 250, 500, 750, 1000, 1250, 1500 or more nucleotides in length and sufficient to specifically hybridize under stringent conditions to mRNA or DNA.

Fragments of the polynucleotides described herein arc also useful to synthesize larger fragments or full-length polynucleotides described herein. For example, a fragment can be hybridized to any portion of an mRNA and a larger or full-length cDNA can be produced.

The fragments are also useful to synthesize antisense molecules of desired length and sequence.

Antisense nucleic acids of the invention can be designed using the nucleotide sequences of SEQ ID NO:1 or SEQ ID NO:3, and constructed using chemical synthesis and enzymatic ligation reactions using procedures known in the art. For example, an antisense nucleic acid (e.g., an antisense oligonucleotide) can be chemically synthesized using naturally occurring nucleotides or variously modified nucleotides designed to increase the biological stability of the molecules or to increase the physical stability of the duplex formed between the antisense and sense nucleic acids, e.g., phosphorothioate derivatives and acridine substituted nucleotides can be used. Examples of modified nucleotides which can be used to generate the antisense nucleic acid include 5-fluorouracil, 5-bromouracil, 5-chlorouracil, 5-iodouracil, hypoxanthine, xanthine, 4-acetylcytosine, 5-(carboxyhydroxylmethyl) uracil, 5-carboxymethylaminomethyl-2-thiouridine, 5-carboxymethylaminomethyluracil, dihydrouracil, beta-D-galactosylqueosine, inosine, N6-isopentenyladenine, 1-methylguanine, 1-methylinosine, 2,2-dimethylguanine, 2-methyladenine, 2-

methylguanine, 3-methylcytosine, 5-methylcytosine, N6-adenine, 7-methylguanine, 5-methylaminomethyluracil, 5-methoxyaminomethyl-2-thiouracil, beta-D-mannosylqueosine, 5'-methoxycarboxymethyluracil, 5-methoxyuracil, 2-methylthio-N6-isopentenyladenine, uracil-5-oxyacetic acid (v), wybutoxosine, pseudouracil, queosine, 2-thiocytosine, 5-methyl-2-thiouracil, 2-thiouracil, 4-thiouracil, 5-methyluracil, uracil-5-oxyacetic acid methylester, uracil-5-oxyacetic acid (v), 5-methyl-2-thiouracil, 3-(3-amino-3-N-2-carboxypropyl) uracil, (acp3)w, and 2,6-diaminopurine. Alternatively, the antisense nucleic acid can be produced biologically using an expression vector into which a nucleic acid has been subcloned in an antisense orientation (i.e., RNA transcribed from the inserted nucleic acid will be of an antisense orientation to a target nucleic acid of interest.

Additionally, the nucleic acid molecules of the invention can be modified at the base moiety, sugar moiety or phosphate backbone to improve, e.g., the stability, hybridization, or solubility of the molecule. For example, the deoxyribose phosphate backbone of the nucleic acids can be modified to generate peptide nucleic acids (see Hyrup *et al.* (1996) *Bioorganic & Medicinal Chemistry 4*:5). As used herein, the terms "peptide nucleic acids" or "PNAs" refer to nucleic acid mimics, e.g., DNA mimics, in which the deoxyribose phosphate backbone is replaced by a pseudopeptide backbone and only the four natural nucleobases are retained. The neutral backbone of PNAs has been shown to allow for specific hybridization to DNA and RNA under conditions of low ionic strength. The synthesis of PNA oligomers can be performed using standard solid phase peptide synthesis protocols as described in Hyrup *et al.* (1996), *supra*; Perry-O'Keefe *et al.* (1996) *Proc. Natl. Acad. Sci. USA 93*:14670. PNAs can be further modified, e.g., to enhance their stability, specificity or cellular uptake, by attaching lipophilic or other helper groups to PNA, by the formation of PNA-DNA chimeras, or by the use of liposomes or other techniques of drug delivery known in the art. The synthesis of PNA-DNA chimeras can be performed as described in Hyrup (1996), *supra*, Finn *et al.* (1996) *Nucleic Acids Res. 24*(17):3357-63, Mag *et al.* (1989) *Nucleic Acids Res. 17*:5973, and Peterser *et al.* (1975) *Bioorganic Med. Chem. Lett. 5*:1119.

The nucleic acid molecules and fragments of the invention can also include other appended groups such as peptides (e.g., for targeting host cell 19459 proteins *in vivo*), or agents facilitating transport across the cell membrane (see, e.g., Letsinger *et*

al. (1989) *Proc. Natl. Acad. Sci. USA 86*:6553-6556; Lemaitre *et al.* (1987) *Proc. Natl. Acad. Sci. USA 84*:648-652; PCT Publication No. WO 88/0918) or the blood brain barrier (see, e.g., PCT Publication No. WO 89/10134). In addition, oligonucleotides can be modified with hybridization-triggered cleavage agents (see, e.g., Krol *et al.* (1988) *Bio-Techniques 6*:958-976) or intercalating agents (see, e.g., Zon (1988) *Pharm Res. 5*:539-549).

The polynucleotides are also useful as primers for PCR to amplify any given region of a 19459 polynucleotide.

The polynucleotides are also useful for constructing recombinant vectors. Such vectors include expression vectors that express a portion of, or all of, the 19459 polypeptides. Vectors also include insertion vectors, used to integrate into another polynucleotide sequence, such as into the cellular genome, to alter *in situ* expression of 19459 genes and gene products. For example, an endogenous 19459 coding sequence can be replaced via homologous recombination with all or part of the coding region containing one or more specifically introduced mutations.

The polynucleotides are also useful for expressing antigenic portions of the 19459 proteins.

The polynucleotides are also useful as probes for determining the chromosomal positions of the 19459 polynucleotides by means of *in situ* hybridization methods, such as FISH (For a review of this technique, see Verma *et al.* (1988) *Human Chromosomes: A Manual of Basic Techniques* (Pergamon Press, New York)), and PCR mapping of somatic cell hybrids. The mapping of the sequences to chromosomes is an important first step in correlating these sequences with genes associated with disease.

Reagents for chromosome mapping can be used individually to mark a single chromosome or a single site on that chromosome, or panels of reagents can be used for marking multiple sites and/or multiple chromosomes. Reagents corresponding to noncoding regions of the genes actually are preferred for mapping purposes. Coding sequences are more likely to be conserved within gene families, thus increasing the chance of cross hybridizations during chromosomal mapping.

Once a sequence has been mapped to a precise chromosomal location, the physical position of the sequence on the chromosome can be correlated with genetic map data. (Such data are found, for example, in *Mendelian Inheritance in Man*, V.

McKusick, available on-line through Johns Hopkins University Welch Medical Library). The relationship between a gene and a disease, mapped to the same chromosomal region, can then be identified through linkage analysis (co-inheritance of physically adjacent genes), described in, for example, Egeland *et al.* (1987) *Nature* *325*:783-787.

Moreover, differences in the DNA sequences between individuals affected and unaffected with a disease associated with a specified gene, can be determined. If a mutation is observed in some or all of the affected individuals but not in any unaffected individuals, then the mutation is likely to be the causative agent of the particular disease. Comparison of affected and unaffected individuals generally involves first looking for structural alterations in the chromosomes, such as deletions or translocations that are visible form chromosome spreads or detectable using PCR based on that DNA sequence. Ultimately, complete sequencing of genes from several individuals can be performed to confirm the presence of a mutation and to distinguish mutations from polymorphisms.

The polynucleotide probes are also useful to determine patterns of the presence of the gene encoding the proteins and their variants with respect to tissue distribution, for example, whether gene duplication has occurred and whether the duplication occurs in all or only a subset of tissues. The genes can be naturally occurring or can have been introduced into a cell, tissue, or organism exogenously.

The polynucleotides are also useful for designing ribozymes corresponding to all, or a part, of the mRNA produced from genes encoding the polynucleotides described herein.

The polynucleotides are also useful for constructing host cells expressing a part, or all, of the polynucleotides and polypeptides.

The polynucleotides are also useful for constructing transgenic animals expressing all, or a part, of the polynucleotides and polypeptides.

The polynucleotides are also useful for making vectors that express part, or all, of the polypeptides.

The polynucleotides are also useful as hybridization probes for determining the level of 19459 nucleic acid expression. Accordingly, the probes can be used to detect the presence of, or to determine levels of, 19459 nucleic acid in cells, tissues, and in organisms. The nucleic acid whose level is determined can be DNA or RNA.

Accordingly, probes corresponding to the polypeptides described herein can be used to assess gene copy number in a given cell, tissue, or organism. This is particularly relevant in cases in which there has been an amplification of the 19459 genes.

Alternatively, the probe can be used in an *in situ* hybridization context to assess the position of extra copies of the 19459 genes, as on extrachromosomal elements or as integrated into chromosomes in which the gene is not normally found, for example as a homogeneously staining region.

These uses are relevant for diagnosis of disorders involving an increase or decrease in expression relative to normal, such as a proliferative disorder, a differentiative or developmental disorder, or a hematopoietic disorder, especially as disclosed hereinabove.

Thus, the present invention provides a method for identifying a disease or disorder associated with aberrant expression or activity of 19459 nucleic acid, in which a test sample is obtained from a subject and nucleic acid (e.g., mRNA, genomic DNA) is detected, wherein the presence of the nucleic acid is diagnostic for a subject having or at risk of developing a disease or disorder associated with aberrant expression or activity of the nucleic acid.

One aspect of the invention relates to diagnostic assays for determining nucleic acid expression as well as activity in the context of a biological sample (e.g., blood, serum, cells, tissue) to determine whether an individual has a disease or disorder, or is at risk of developing a disease or disorder, associated with aberrant nucleic acid expression or activity. Such assays can be used for prognostic or predictive purpose to thereby prophylactically treat an individual prior to the onset of a disorder characterized by or associated with expression or activity of the nucleic acid molecules.

In vitro techniques for detection of mRNA include Northern hybridizations and *in situ* hybridizations. *In vitro* techniques for detecting DNA includes Southern hybridizations and *in situ* hybridization.

Probes can be used as a part of a diagnostic test kit for identifying cells or tissues that express a 19459 protein, such as by measuring a level of a 19459 protein-encoding nucleic acid in a sample of cells from a subject e.g., mRNA or genomic DNA, or determining if a 19459 gene has been mutated.

Nucleic acid expression assays are useful for drug screening to identify compounds that modulate 19459 nucleic acid expression (e.g., antisense, polypeptides, peptidomimetics, small molecules or other drugs). A cell is contacted with a candidate compound and the expression of mRNA determined. The level of expression of 19459 mRNA in the presence of the candidate compound is compared to the level of expression of 19459 mRNA in the absence of the candidate compound. The candidate compound can then be identified as a modulator of nucleic acid expression based on this comparison and be used, for example to treat a disorder characterized by aberrant nucleic acid expression. The modulator can bind to the nucleic acid or indirectly modulate expression, such as by interacting with other cellular components that affect nucleic acid expression.

Modulatory methods can be performed *in vitro* (e.g., by culturing the cell with the agent) or, alternatively, *in vivo* (e.g., by administering the agent to a subject) in patients or in transgenic animals.

The invention thus provides a method for identifying a compound that can be used to treat a disorder associated with nucleic acid expression of the 19459 gene. The method typically includes assaying the ability of the compound to modulate the expression of the 19459 nucleic acid and thus identifying a compound that can be used to treat a disorder characterized by undesired 19459 nucleic acid expression.

The assays can be performed in cell-based and cell-free systems. Cell-based assays include cells naturally expressing the 19459 nucleic acid, such as discussed hereinabove, or recombinant cells genetically engineered to express specific nucleic acid sequences.

Alternatively, candidate compounds can be assayed *in vivo* in patients or in transgenic animals.

The assay for 19459 nucleic acid expression can involve direct assay of nucleic acid levels, such as mRNA levels, or on collateral compounds involved in the signal pathway (such as cyclic AMP or phosphatidylinositol turnover). Further, the expression of genes that are up- or down-regulated in response to the receptor protein signal pathway can also be assayed. In this embodiment the regulatory regions of these genes can be operably linked to a reporter gene such as luciferase.

Thus, modulators of 19459 gene expression can be identified in a method wherein a cell is contacted with a candidate compound and the expression of mRNA

determined. The level of expression of mRNA in the presence of the candidate compound is compared to the level of expression of mRNA in the absence of the candidate compound. The candidate compound can then be identified as a modulator of nucleic acid expression based on this comparison and be used, for example to treat a disorder characterized by aberrant nucleic acid expression. When expression of mRNA is statistically significantly greater in the presence of the candidate compound than in its absence, the candidate compound is identified as a stimulator of nucleic acid expression. When nucleic acid expression is statistically significantly less in the presence of the candidate compound than in its absence, the candidate compound is identified as an inhibitor of nucleic acid expression.

Accordingly, the invention provides methods of treatment, with the nucleic acid as a target, using a compound identified through drug screening as a gene modulator to modulate nucleic acid expression. Modulation includes both up-regulation (i.e. activation or agonization) or down-regulation (suppression or antagonization) or effects on nucleic acid activity (e.g. when nucleic acid is mutated or improperly modified) Treatment is of disorders characterized by aberrant expression or activity of the nucleic acid.

Treatment is defined as the application or administration of a therapeutic agent to a patient, or application or administration of a therapeutic agent to an isolated tissue or cell line from a patient, who has a disease, a symptom of disease or a predisposition toward a disease, with the purpose to cure, heal, alleviate, relieve, alter, remedy, ameliorate, improve or affect the disease, the symptoms of disease or the predisposition toward disease. "Subject", as used herein, can refer to a mammal, e.g. a human, or to an experimental or animal or disease model. The subject can also be a non-human animal, e.g. a horse, cow, goat, or other domestic animal. A therapeutic agent includes, but is not limited to, small molecules, peptides, antibodies, ribozymes and antisense oligonucleotides.

"Misexpression or aberrant expression", as used herein, refers to a non-wild type pattern of gene expression, at the RNA or protein level. It includes: expression at non-wild type levels, i.e., over or under expression; a pattern of expression that differs from wild type in terms of the time or stage at which the gene is expressed, e.g., increased or decreased expression (as compared with wild type) at a predetermined developmental period or stage; a pattern of expression that differs from wild type in terms of decreased

expression (as compared with wild type) in a predetermined cell type or tissue type; a pattern of expression that differs from wild type in terms of the splicing size, amino acid sequence, post-transitional modification, or biological activity of the expressed polypeptide; a pattern of expression that differs from wild type in terms of the effect of an environmental stimulus or extracellular stimulus on expression of the gene, e.g., a pattern of increased or decreased expression (as compared with wild type) in the presence of an increase or decrease in the strength of the stimulus.

A modulator for nucleic acid expression can be a small molecule or drug identified using the screening assays described herein as long as the drug or small molecule inhibits the nucleic acid expression.

The polynucleotides are also useful for monitoring the effectiveness of modulating compounds on the expression or activity of the gene in clinical trials or in a treatment regimen. Thus, the gene expression pattern can serve as a barometer for the continuing effectiveness of treatment with the compound, particularly with compounds to which a patient can develop resistance. The gene expression pattern can also serve as a marker indicative of a physiological response of the affected cells to the compound. Accordingly, such monitoring would allow either increased administration of the compound or the administration of alternative compounds to which the patient has not become resistant. Similarly, if the level of nucleic acid expression falls below a desirable level, administration of the compound could be commensurately decreased.

Monitoring can be, for example, as follows: (i) obtaining a pre-administration sample from a subject prior to administration of the agent; (ii) detecting the level of expression of a specified mRNA or genomic DNA of the invention in the pre-administration sample; (iii) obtaining one or more post-administration samples from the subject; (iv) detecting the level of expression or activity of the mRNA or genomic DNA in the post-administration samples; (v) comparing the level of expression or activity of the mRNA or genomic DNA in the pre-administration sample with the mRNA or genomic DNA in the post-administration sample or samples; and (vi) increasing or decreasing the administration of the agent to the subject accordingly.

The polynucleotides are also useful in diagnostic assays for qualitative changes in 19459 nucleic acid, and particularly in qualitative changes that lead to pathology. The polynucleotides can be used to detect mutations in 19459 genes and gene expression products such as mRNA. The polynucleotides can be used as hybridization probes to

detect naturally-occurring genetic mutations in the 19459 gene and thereby to determine whether a subject with the mutation is at risk for a disorder caused by the mutation. Mutations include deletion, addition, or substitution of one or more nucleotides in the gene, chromosomal rearrangement, such as inversion or transposition, modification of genomic DNA, such as aberrant methylation patterns or changes in gene copy number, such as amplification. Detection of a mutated form of the gene associated with a dysfunction provides a diagnostic tool for an active disease or susceptibility to disease when the disease results from overexpression, underexpression, or altered expression of a 19459 protein.

Mutations in the gene can be detected at the nucleic acid level by a variety of techniques. Genomic DNA can be analyzed directly or can be amplified by using PCR prior to analysis. RNA or cDNA can be used in the same way.

In certain embodiments, detection of the mutation involves the use of a probe/primer in a polymerase chain reaction (PCR) (see, e.g. U.S. Patent Nos. 4,683,195 and 4,683,202), such as anchor PCR or RACE PCR, or, alternatively, in a ligation chain reaction (LCR) (see, e.g., Landegran *et al.*, *Science 241*:1077-1080 (1988); and Nakazawa *et al.*, *PNAS 91*:360-364 (1994)), the latter of which can be particularly useful for detecting point mutations in the gene (see Abravaya *et al.*, *Nucleic Acids Res. 23*:675-682 (1995)). This method can include the steps of collecting a sample of cells from a patient, isolating nucleic acid (e.g., genomic, mRNA or both) from the cells of the sample, contacting the nucleic acid sample with one or more primers which specifically hybridize to a gene under conditions such that hybridization and amplification of the gene (if present) occurs, and detecting the presence or absence of an amplification product, or detecting the size of the amplification product and comparing the length to a control sample. Deletions and insertions can be detected by a change in size of the amplified product compared to the normal genotype. Point mutations can be identified by hybridizing amplified DNA to normal RNA or antisense DNA sequences.

It is anticipated that PCR and/or LCR may be desirable to use as a preliminary amplification step in conjunction with any of the techniques used for detecting mutations described herein.

Alternative amplification methods include: self sustained sequence replication (Guatelli *et al.* (1990) *Proc. Natl. Acad. Sci. USA 87*:1874-1878), transcriptional amplification system (Kwoh *et al.* (1989) *Proc. Natl. Acad. Sci. USA 86*:1173-1177),

Q-Beta Replicase (Lizardi *et al.* (1988) *Bio/Technology 6*:1197), or any other nucleic acid amplification method, followed by the detection of the amplified molecules using techniques well-known to those of skill in the art. These detection schemes are especially useful for the detection of nucleic acid molecules if such molecules are present in very low numbers.

Alternatively, mutations in a 19459 gene can be directly identified, for example, by alterations in restriction enzyme digestion patterns determined by gel electrophoresis.

Further, sequence-specific ribozymes (U.S.Patent No. 5,498,531) can be used to score for the presence of specific mutations by development or loss of a ribozyme cleavage site.

Perfectly matched sequences can be distinguished from mismatched sequences by nuclease cleavage digestion assays or by differences in melting temperature.

Sequence changes at specific locations can also be assessed by nuclease protection assays such as RNase and S1 protection or the chemical cleavage method.

Furthermore, sequence differences between a mutant 19459 gene and a wild-type gene can be determined by direct DNA sequencing. A variety of automated sequencing procedures can be utilized when performing the diagnostic assays ((1995) *Biotechniques 19*:448), including sequencing by mass spectrometry (see, e.g., PCT International Publication No. WO 94/16101; Cohen *et al.*, *Adv. Chromatogr. 36*:127-162 (1996); and Griffin *et al.*, *Appl. Biochem. Biotechnol. 38*:147-159 (1993)).

Other methods for detecting mutations in the gene include methods in which protection from cleavage agents is used to detect mismatched bases in RNA/RNA or RNA/DNA duplexes (Myers *et al.*, *Science 230*:1242 (1985)); Cotton *et al.*, *PNAS 85*:4397 (1988); Saleeba *et al.*, *Meth. Enzymol. 217*:286-295 (1992)), electrophoretic mobility of mutant and wild type nucleic acid is compared (Orita *et al.*, *PNAS 86*:2766 (1989); Cotton *et al.*, *Mutat. Res. 285*:125-144 (1993); and Hayashi *et al.*, *Genet. Anal. Tech. Appl. 9*:73-79 (1992)), and movement of mutant or wild-type fragments in polyacrylamide gels containing a gradient of denaturant is assayed using denaturing gradient gel electrophoresis (Myers *et al.*, *Nature 313*:495 (1985)). The sensitivity of the assay may be enhanced by using RNA (rather than DNA), in which the secondary structure is more sensitive to a change in sequence. In one embodiment, the subject method utilizes heteroduplex analysis to separate double stranded heteroduplex molecules on the basis of changes in electrophoretic mobility (Keen *et al.* (1991)

Trends Genet. 7:5). Examples of other techniques for detecting point mutations include, selective oligonucleotide hybridization, selective amplification, and selective primer extension.

In other embodiments, genetic mutations can be identified by hybridizing a sample and control nucleic acids, e.g., DNA or RNA, to high density arrays containing hundreds or thousands of oligonucleotide probes (Cronin *et al.* (1996) *Human Mutation* 7:244-255; Kozal *et al.* (1996) *Nature Medicine* 2:753-759). For example, genetic mutations can be identified in two dimensional arrays containing light-generated DNA probes as described in Cronin *et al. supra.* Briefly, a first hybridization array of probes can be used to scan through long stretches of DNA in a sample and control to identify base changes between the sequences by making linear arrays of sequential overlapping probes. This step allows the identification of point mutations. This step is followed by a second hybridization array that allows the characterization of specific mutations by using smaller, specialized probe arrays complementary to all variants or mutations detected. Each mutation array is composed of parallel probe sets, one complementary to the wild-type gene and the other complementary to the mutant gene.

The polynucleotides are also useful for testing an individual for a genotype that while not necessarily causing the disease, nevertheless affects the treatment modality. Thus, the polynucleotides can be used to study the relationship between an individual's genotype and the individual's response to a compound used for treatment (pharmacogenomic relationship). In the present case, for example, a mutation in the gene that results in altered affinity for ligand could result in an excessive or decreased drug effect with standard concentrations of ligand that activates the protein. Accordingly, the polynucleotides described herein can be used to assess the mutation content of the gene in an individual in order to select an appropriate compound or dosage regimen for treatment.

Thus polynucleotides displaying genetic variations that affect treatment provide a diagnostic target that can be used to tailor treatment in an individual. Accordingly, the production of recombinant cells and animals containing these polymorphisms allow effective clinical design of treatment compounds and dosage regimens.

The methods can involve obtaining a control biological sample from a control subject, contacting the control sample with a compound or agent capable of detecting

mRNA, or genomic DNA, such that the presence of mRNA or genomic DNA is detected in the biological sample, and comparing the presence of mRNA or genomic DNA in the control sample with the presence of mRNA or genomic DNA in the test sample.

The polynucleotides are also useful for chromosome identification when the sequence is identified with an individual chromosome and to a particular location on the chromosome. First, the DNA sequence is matched to the chromosome by *in situ* or other chromosome-specific hybridization. Sequences can also be correlated to specific chromosomes by preparing PCR primers that can be used for PCR screening of somatic cell hybrids containing individual chromosomes from the desired species. Only hybrids containing the chromosome containing the gene homologous to the primer will yield an amplified fragment. Sublocalization can be achieved using chromosomal fragments. Other strategies include prescreening with labeled flow-sorted chromosomes and preselection by hybridization to chromosome-specific libraries. Further mapping strategies include fluorescence *in situ* hybridization which allows hybridization with probes shorter than those traditionally used. Reagents for chromosome mapping can be used individually to mark a single chromosome or a single site on the chromosome, or panels of reagents can be used for marking multiple sites and/or multiple chromosomes. Reagents corresponding to noncoding regions of the genes actually are preferred for mapping purposes. Coding sequences are more likely to be conserved within gene families, thus increasing the chance of cross hybridizations during chromosomal mapping.

The polynucleotides can also be used to identify individuals from small biological samples. This can be done for example using restriction fragment-length polymorphism (RFLP) to identify an individual. Thus, the polynucleotides described herein are useful as DNA markers for RFLP (See U.S. Patent No. 5,272,057).

Furthermore, the sequence can be used to provide an alternative technique which determines the actual DNA sequence of selected fragments in the genome of an individual. Thus, the sequences described herein can be used to prepare two PCR primers from the 5' and 3' ends of the sequences. These primers can then be used to amplify DNA from an individual for subsequent sequencing.

Panels of corresponding DNA sequences from individuals prepared in this manner can provide unique individual identifications, as each individual will have a

unique set of such DNA sequences. It is estimated that allelic variation in humans occurs with a frequency of about once per each 500 bases. Allelic variation occurs to some degree in the coding regions of these sequences, and to a greater degree in the noncoding regions. The sequences can be used to obtain such identification sequences from individuals and from tissue. The sequences represent unique fragments of the human genome. Each of the sequences described herein can, to some degree, be used as a standard against which DNA from an individual can be compared for identification purposes.

If a panel of reagents from the sequences is used to generate a unique identification database for an individual, those same reagents can later be used to identify tissue from that individual. Using the unique identification database, positive identification of the individual, living or dead, can be made from extremely small tissue samples.

The polynucleotides can also be used in forensic identification procedures. PCR technology can be used to amplify DNA sequences taken from very small biological samples, such as a single hair follicle, body fluids (eg. blood, saliva, or semen). The amplified sequence can then be compared to a standard allowing identification of the origin of the sample.

The polynucleotides can thus be used to provide polynucleotide reagents, e.g., PCR primers, targeted to specific loci in the human genome, which can enhance the reliability of DNA-based forensic identifications by, for example, providing another "identification marker" (i.e. another DNA sequence that is unique to a particular individual). As described above, actual base sequence information can be used for identification as an accurate alternative to patterns formed by restriction enzyme generated fragments. Sequences targeted to the noncoding region are particularly useful since greater polymorphism occurs in the noncoding regions, making it easier to differentiate individuals using this technique.

The polynucleotides can further be used to provide polynucleotide reagents, e.g., labeled or labelable probes which can be used in, for example, an *in situ* hybridization technique, to identify a specific tissue. This is useful in cases in which a forensic pathologist is presented with a tissue of unknown origin. Panels of probes can be used to identify tissue by species and/or by organ type.

In a similar fashion, these primers and probes can be used to screen tissue culture for contamination (i.e. screen for the presence of a mixture of different types of cells in a culture).

Alternatively, the polynucleotides can be used directly to block transcription or translation of 19459 gene sequences by means of antisense or ribozyme constructs. Thus, in a disorder characterized by abnormally high or undesirable 19459 gene expression, nucleic acids can be directly used for treatment.

The polynucleotides are thus useful as antisense constructs to control 19459 gene expression in cells, tissues, and organisms. A DNA antisense polynucleotide is designed to be complementary to a region of the gene involved in transcription, preventing transcription and hence production of protein. An antisense RNA or DNA polynucleotide would hybridize to the mRNA and thus block translation of mRNA into protein.

Examples of antisense molecules useful to inhibit nucleic acid expression include antisense molecules complementary to a fragment of the 5' untranslated region of SEQ ID NO:1 which also includes the start codon and antisense molecules which are complementary to a fragment of the 3' untranslated region of SEQ ID NO:1.

Alternatively, a class of antisense molecules can be used to inactivate mRNA in order to decrease expression of 19459 nucleic acid. Accordingly, these molecules can treat a disorder characterized by abnormal or undesired 19459 nucleic acid expression. This technique involves cleavage by means of ribozymes containing nucleotide sequences complementary to one or more regions in the mRNA that attenuate the ability of the mRNA to be translated. Possible regions include coding regions and particularly coding regions corresponding to the catalytic and other functional activities of the protein, such as ligand binding. It is understood that these regions include any of those specific domains, sites, segments, loops, and the like that are disclosed as specific regions or sites herein.

The polynucleotides also provide vectors for gene therapy in patients containing cells that are aberrant in 19459 gene expression. Thus, recombinant cells, which include the patient's cells that have been engineered *ex vivo* and returned to the patient, are introduced into an individual where the cells produce the desired protein to treat the individual.

The invention also encompasses kits for detecting the presence of a 19459 nucleic acid in a biological sample. For example, the kit can comprise reagents such as a labeled or labelable nucleic acid or agent capable of detecting 19459 nucleic acid in a biological sample; means for determining the amount of 19459 nucleic acid in the sample; and means for comparing the amount of 19459 nucleic acid in the sample with a standard. The compound or agent can be packaged in a suitable container. The kit can further comprise instructions for using the kit to detect 19459 mRNA or DNA.

Computer Readable Means

The nucleotide or amino acid sequences of the invention are also provided in a variety of mediums to facilitate use thereof. As used herein, "provided" refers to a manufacture, other than an isolated nucleic acid or amino acid molecule, which contains a nucleotide or amino acid sequence of the present invention. Such a manufacture provides the nucleotide or amino acid sequences, or a subset thereof (e.g., a subset of open reading frames (ORFs)) in a form which allows a skilled artisan to examine the manufacture using means not directly applicable to examining the nucleotide or amino acid sequences, or a subset thereof, as they exists in nature or in purified form.

In one application of this embodiment, a nucleotide or amino acid sequence of the present invention can be recorded on computer readable media. As used herein, "computer readable media" refers to any medium that can be read and accessed directly by a computer. Such media include, but are not limited to: magnetic storage media, such as floppy discs, hard disc storage medium, and magnetic tape; optical storage media such as CD-ROM; electrical storage media such as RAM and ROM; and hybrids of these categories such as magnetic/optical storage media. The skilled artisan will readily appreciate how any of the presently known computer readable mediums can be used to create a manufacture comprising computer readable medium having recorded thereon a nucleotide or amino acid sequence of the present invention.

As used herein, "recorded" refers to a process for storing information on computer readable medium. The skilled artisan can readily adopt any of the presently known methods for recording information on computer readable medium to generate manufactures comprising the nucleotide or amino acid sequence information of the present invention.

A variety of data storage structures are available to a skilled artisan for creating a computer readable medium having recorded thereon a nucleotide or amino acid sequence of the present invention. The choice of the data storage structure will generally be based on the means chosen to access the stored information. In addition, a variety of data processor programs and formats can be used to store the nucleotide sequence information of the present invention on computer readable medium. The sequence information can be represented in a word processing text file, formatted in commercially-available software such as WordPerfect and MicroSoft Word, or represented in the form of an ASCII file, stored in a database application, such as DB2, Sybase, Oracle, or the like. The skilled artisan can readily adapt any number of dataprocessor structuring formats (e.g., text file or database) in order to obtain computer readable medium having recorded thereon the nucleotide sequence information of the present invention.

By providing the nucleotide or amino acid sequences of the invention in computer readable form, the skilled artisan can routinely access the sequence information for a variety of purposes. For example, one skilled in the art can use the nucleotide or amino acid sequences of the invention in computer readable form to compare a target sequence or target structural motif with the sequence information stored within the data storage means. Search means are used to identify fragments or regions of the sequences of the invention which match a particular target sequence or target motif.

As used herein, a "target sequence" can be any DNA or amino acid sequence of six or more nucleotides or two or more amino acids. A skilled artisan can readily recognize that the longer a target sequence is, the less likely a target sequence will be present as a random occurrence in the database. The most preferred sequence length of a target sequence is from about 10 to 100 amino acids or from about 30 to 300 nucleotide residues. However, it is well recognized that commercially important fragments, such as sequence fragments involved in gene expression and protein processing, may be of shorter length.

As used herein, "a target structural motif," or "target motif," refers to any rationally selected sequence or combination of sequences in which the sequence(s) are chosen based on a three-dimensional configuration which is formed upon the folding of the target motif. There are a variety of target motifs known in the art. Protein

target motifs include, but are not limited to, enzyme active sites and signal sequences. Nucleic acid target motifs include, but are not limited to, promoter sequences, hairpin structures and inducible expression elements (protein binding sequences).

Computer software is publicly available which allows a skilled artisan to access sequence information provided in a computer readable medium for analysis and comparison to other sequences. A variety of known algorithms are disclosed publicly and a variety of commercially available software for conducting search means are and can be used in the computer-based systems of the present invention. Examples of such software includes, but is not limited to, MacPattern (EMBL), BLASTN and BLASTX (NCBIA).

For example, software which implements the BLAST (Altschul *et al.* (1990) *J. Mol. Biol. 215*:403-410) and BLAZE (Brutlag *et al.* (1993) *Comp. Chem. 17*:203-207) search algorithms on a Sybase system can be used to identify open reading frames (ORFs) of the sequences of the invention which contain homology to ORFs or proteins from other libraries. Such ORFs are protein encoding fragments and are useful in producing commercially important proteins such as enzymes used in various reactions and in the production of commercially useful metabolites.

Vectors/host cells

The invention also provides vectors containing the 19459 polynucleotides. The term "vector" refers to a vehicle, preferably a nucleic acid molecule, that can transport the polynucleotides. When the vector is a nucleic acid molecule, the polynucleotides are covalently linked to the vector nucleic acid. With this aspect of the invention, the vector includes a plasmid, single or double stranded phage, a single or double stranded RNA or DNA viral vector, or artificial chromosome, such as a BAC, PAC, YAC, OR MAC.

A vector can be maintained in the host cell as an extrachromosomal element where it replicates and produces additional copies of the polynucleotides. Alternatively, the vector may integrate into the host cell genome and produce additional copies of the polynucleotides when the host cell replicates.

The invention provides vectors for the maintenance (cloning vectors) or vectors for expression (expression vectors) of the polynucleotides. The vectors can function in procaryotic or eukaryotic cells or in both (shuttle vectors).

Expression vectors contain cis-acting regulatory regions that are operably linked in the vector to the polynucleotides such that transcription of the polynucleotides is allowed in a host cell. The polynucleotides can be introduced into the host cell with a separate polynucleotide capable of affecting transcription. Thus, the second polynucleotide may provide a trans-acting factor interacting with the cis-regulatory control region to allow transcription of the polynucleotides from the vector. Alternatively, a trans-acting factor may be supplied by the host cell. Finally, a trans-acting factor can be produced from the vector itself.

It is understood, however, that in some embodiments, transcription and/or translation of the polynucleotides can occur in a cell-free system.

The regulatory sequence to which the polynucleotides described herein can be operably linked include promoters for directing mRNA transcription. These include, but are not limited to, the left promoter from bacteriophage λ, the lac, TRP, and TAC promoters from *E. coli*, the early and late promoters from SV40, the CMV immediate early promoter, the adenovirus early and late promoters, and retrovirus long-terminal repeats.

In addition to control regions that promote transcription, expression vectors may also include regions that modulate transcription, such as repressor binding sites and enhancers. Examples include the SV40 enhancer, the cytomegalovirus immediate early enhancer, polyoma enhancer, adenovirus enhancers, and retrovirus LTR enhancers.

In addition to containing sites for transcription initiation and control, expression vectors can also contain sequences necessary for transcription termination and, in the transcribed region a ribosome binding site for translation. Other regulatory control elements for expression include initiation and termination codons as well as polyadenylation signals. The person of ordinary skill in the art would be aware of the numerous regulatory sequences that are useful in expression vectors. Such regulatory sequences are described, for example, in Sambrook *et al.*, *Molecular Cloning: A Laboratory Manual. 2nd. ed.*, Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY, (1989).

A variety of expression vectors can be used to express a 19459 polynucleotide. Such vectors include chromosomal, episomal, and virus-derived vectors, for example vectors derived from bacterial plasmids, from bacteriophage, from yeast episomes, from yeast chromosomal elements, including yeast artificial chromosomes, from viruses such

as baculoviruses, papovaviruses such as SV40, Vaccinia viruses, adenoviruses, poxviruses, pseudorabies viruses, and retroviruses. Vectors may also be derived from combinations of these sources such as those derived from plasmid and bacteriophage genetic elements, eg. cosmids and phagemids. Appropriate cloning and expression vectors for prokaryotic and eukaryotic hosts are described in Sambrook *et al.*, *Molecular Cloning: A Laboratory Manual. 2nd. ed.*, Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY, (1989).

The regulatory sequence may provide constitutive expression in one or more host cells (i.e. tissue specific) or may provide for inducible expression in one or more cell types such as by temperature, nutrient additive, or exogenous factor such as a hormone or other ligand. A variety of vectors providing for constitutive and inducible expression in prokaryotic and eukaryotic hosts are well known to those of ordinary skill in the art.

The polynucleotides can be inserted into the vector nucleic acid by well-known methodology. Generally, the DNA sequence that will ultimately be expressed is joined to an expression vector by cleaving the DNA sequence and the expression vector with one or more restriction enzymes and then ligating the fragments together. Procedures for restriction enzyme digestion and ligation are well known to those of ordinary skill in the art.

The vector containing the appropriate polynucleotide can be introduced into an appropriate host cell for propagation or expression using well-known techniques. Bacterial cells include, but are not limited to, *E. coli*, *Streptomyces*, and *Salmonella typhimurium*. Eukaryotic cells include, but are not limited to, yeast, insect cells such as *Drosophila*, animal cells such as COS and CHO cells, and plant cells.

It is further recognized that the nucleic acid sequences of the invention can be altered to contain codons, which are preferred, or non preferred, for a particular expression system. For example, the nucleic acid can be one in which at least one altered codon, and preferably at least 10%, or 20% of the codons have been altered such that the sequence is optimized for expression in E. coli, yeast, human, insect, or CHO cells. Methods for determining such codon usage are well known in the art.

As described herein, it may be desirable to express the polypeptide as a fusion protein. Accordingly, the invention provides fusion vectors that allow for the production of the polypeptides. Fusion vectors can increase the expression of a recombinant protein, increase the solubility of the recombinant protein, and aid in the purification of

the protein by acting for example as a ligand for affinity purification. A proteolytic cleavage site may be introduced at the junction of the fusion moiety so that the desired polypeptide can ultimately be separated from the fusion moiety. Proteolytic enzymes include, but are not limited to, factor Xa, thrombin, and enterokinase. Typical fusion expression vectors include pGEX (Smith *et al.*, *Gene 67*:31-40 (1988)), pMAL (New England Biolabs, Beverly, MA) and pRIT5 (Pharmacia, Piscataway, NJ) which fuse glutathione S-transferase (GST), maltose E binding protein, or protein A, respectively, to the target recombinant protein. Examples of suitable inducible non-fusion *E. coli* expression vectors include pTrc (Amann *et al.*, *Gene 69*:301-315 (1988)) and pET 11d (Studier *et al.*, *Gene Expression Technology: Methods in Enzymology 185*:60-89 (1990)).

Recombinant protein expression can be maximized in a host bacteria by providing a genetic background wherein the host cell has an impaired capacity to proteolytically cleave the recombinant protein. (Gottesman, S., *Gene Expression Technology: Methods in Enzymology* 185, Academic Press, San Diego, California (1990) 119-128). Alternatively, the sequence of the polynucleotide of interest can be altered to provide preferential codon usage for a specific host cell, for example *E. coli*. (Wada *et al.*, *Nucleic Acids Res. 20*:2111-2118 (1992)).

The polynucleotides can also be expressed by expression vectors that are operative in yeast. Examples of vectors for expression in yeast e.g., *S. cerevisiae* include pYepSec1 (Baldari, *et al.*, *EMBO J. 6*:229-234 (1987)), pMFa (Kurjan *et al.*, *Cell 30*:933-943(1982)), pJRY88 (Schultz *et al.*, *Gene 54*:113-123 (1987)), and pYES2 (Invitrogen Corporation, San Diego, CA).

The polynucleotides can also be expressed in insect cells using, for example, baculovirus expression vectors. Baculovirus vectors available for expression of proteins in cultured insect cells (e.g., Sf 9 cells) include the pAc series (Smith *et al.*, *Mol. Cell Biol. 3*:2156-2165 (1983)) and the pVL series (Lucklow *et al.*, *Virology 170*:31-39 (1989)).

In certain embodiments of the invention, the polynucleotides described herein are expressed in mammalian cells using mammalian expression vectors. Examples of mammalian expression vectors include pCDM8 (Seed, B. *Nature 329*:840 (1987)) and pMT2PC (Kaufman *et al.*, *EMBO J. 6*:187-195 (1987)).

The expression vectors listed herein are provided by way of example only of the well-known vectors available to those of ordinary skill in the art that would be useful to express the 19459 polynucleotides. The person of ordinary skill in the art would be aware of other vectors suitable for maintenance propagation or expression of the polynucleotides described herein. These are found for example in Sambrook, J., Fritsh, E. F., and Maniatis, T. *Molecular Cloning: A Laboratory Manual. 2nd, ed., Cold Spring Harbor Laboratory,* Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY, 1989.

The invention also encompasses vectors in which the nucleic acid sequences described herein are cloned into the vector in reverse orientation, but operably linked to a regulatory sequence that permits transcription of antisense RNA. Thus, an antisense transcript can be produced to all, or to a portion, of the polynucleotide sequences described herein, including both coding and non-coding regions. Expression of this antisense RNA is subject to each of the parameters described above in relation to expression of the sense RNA (regulatory sequences, constitutive or inducible expression, tissue-specific expression).

The invention also relates to recombinant host cells containing the vectors described herein. Host cells therefore include prokaryotic cells, lower eukaryotic cells such as yeast, other eukaryotic cells such as insect cells, and higher eukaryotic cells such as mammalian cells.

The recombinant host cells are prepared by introducing the vector constructs described herein into the cells by techniques readily available to the person of ordinary skill in the art. These include, but are not limited to, calcium phosphate transfection, DEAE-dextran-mediated transfection, cationic lipid-mediated transfection, electroporation, transduction, infection, lipofection, and other techniques such as those found in Sambrook, *et al.* (*Molecular Cloning: A Laboratory Manual. 2nd, ed., Cold Spring Harbor Laboratory,* Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY, 1989).

Host cells can contain more than one vector. Thus, different nucleotide sequences can be introduced on different vectors of the same cell. Similarly, the polynucleotides can be introduced either alone or with other polynucleotides that are not related to the polynucleotides such as those providing trans-acting factors for expression

vectors. When more than one vector is introduced into a cell, the vectors can be introduced independently, co-introduced or joined to the polynucleotide vector.

In the case of bacteriophage and viral vectors, these can be introduced into cells as packaged or encapsulated virus by standard procedures for infection and transduction. Viral vectors can be replication-competent or replication-defective. In the case in which viral replication is defective, replication will occur in host cells providing functions that complement the defects.

Vectors generally include selectable markers that enable the selection of the subpopulation of cells that contain the recombinant vector constructs. The marker can be contained in the same vector that contains the polynucleotides described herein or may be on a separate vector. Markers include tetracycline or ampicillin-resistance genes for prokaryotic host cells and dihydrofolate reductase or neomycin resistance for eukaryotic host cells. However, any marker that provides selection for a phenotypic trait will be effective.

While the mature proteins can be produced in bacteria, yeast, mammalian cells, and other cells under the control of the appropriate regulatory sequences, cell-free transcription and translation systems can also be used to produce these proteins using RNA derived from the DNA constructs described herein.

Where secretion of the polypeptide is desired, appropriate secretion signals are incorporated into the vector. The signal sequence can be endogenous to the polypeptides or heterologous to these polypeptides.

Where the polypeptide is not secreted into the medium, the protein can be isolated from the host cell by standard disruption procedures, including freeze thaw, sonication, mechanical disruption, use of lysing agents and the like. The polypeptide can then be recovered and purified by well-known purification methods including ammonium sulfate precipitation, acid extraction, anion or cationic exchange chromatography, phosphocellulose chromatography, hydrophobic-interaction chromatography, affinity chromatography, hydroxylapatite chromatography, lectin chromatography, or high performance liquid chromatography.

It is also understood that depending upon the host cell in recombinant production of the polypeptides described herein, the polypeptides can have various glycosylation patterns, depending upon the cell, or maybe non-glycosylated as when produced in

bacteria. In addition, the polypeptides may include an initial modified methionine in some cases as a result of a host-mediated process.

Host cells of particular interest include those derived from the tissues in which the proteins of the invention are expressed, including but not limited to the tissues shown in Figure 5, especially heart and kidney.

Uses of vectors and host cells

It is understood that "host cells" and "recombinant host cells" refer not only to the particular subject cell but also to the progeny or potential progeny of such a cell. Because certain modifications may occur in succeeding generations due to either mutation or environmental influences, such progeny may not, in fact, be identical to the parent cell, but are still included within the scope of the term as used herein.

A "purified preparation of cells", as used herein, refers to, in the case of plant or animal cells, an in vitro preparation of cells and not an entire intact plant or animal. In the case of cultured cells or microbial cells, it consists of a preparation of at least 10% and more preferably 50% of the subject cells.

The host cells expressing the polypeptides described herein, and particularly recombinant host cells, have a variety of uses. First, the cells are useful for producing 19459 proteins or polypeptides that can be further purified to produce desired amounts of the protein or fragments. Thus, host cells containing expression vectors are useful for polypeptide production.

Host cells are also useful for conducting cell-based assays involving the 19459 protein or fragments. Thus, a recombinant host cell expressing the native protein is useful to assay for compounds that stimulate or inhibit 19459 protein function. This can include ligand binding, gene expression at the level of transcription or translation, G-protein interaction, and components of the signal transduction pathway.

Cells of particular relevance are those in which the protein is expressed as disclosed herein.

Host cells are also useful for identifying mutants in which these functions are affected. If the mutants naturally occur and give rise to a pathology, host cells containing the mutations are useful to assay compounds that have a desired effect on the mutant protein (for example, stimulating or inhibiting function) which may not be indicated by their effect on the native protein.

Recombinant host cells are also useful for expressing the chimeric polypeptides described herein to assess compounds that activate or suppress activation by means of a heterologous amino terminal extracellular domain (or other binding region). Alternatively, a heterologous region spanning the entire transmembrane domain (or parts thereof) can be used to assess the effect of a desired amino terminal extracellular domain (or other binding region) on any given host cell. In this embodiment, a region spanning the entire transmembrane domain (or parts thereof) compatible with the specific host cell is used to make the chimeric vector. Alternatively, a heterologous carboxy terminal intracellular, e.g., signal transduction, domain can be introduced into the host cell.

Further, mutant 19459 proteins can be designed in which one or more of the various functions is engineered to be increased or decreased (e.g., ligand binding or G-protein binding) and used to augment or replace 19459 proteins in an individual. Thus, host cells can provide a therapeutic benefit by replacing an aberrant protein or providing an aberrant protein that provides a therapeutic result. In one embodiment, the cells provide 19459 proteins that are abnormally active.

In another embodiment, the cells provide proteins that are abnormally inactive. These can compete with the endogenous protein in the individual.

In another embodiment, cells expressing the proteins that cannot be activated, are introduced into an individual in order to compete with the endogenous protein for ligand. For example, in the case in which excessive ligand is part of a treatment modality, it may be necessary to inactivate this ligand at a specific point in treatment. Providing cells that compete for the ligand, but which cannot be affected by protein activation would be beneficial.

Homologously recombinant host cells can also be produced that allow the *in situ* alteration of the endogenous polynucleotide sequences in a host cell genome. The host cell includes, but is not limited to, a stable cell line, cell *in vivo*, or cloned microorganism. This technology is more fully described in WO 93/09222, WO 91/12650, WO 91/06667, U.S. 5,272,071, and U.S. 5,641,670. Briefly, specific polynucleotide sequences corresponding to the 19459 polynucleotides or sequences proximal or distal to a 19459 gene are allowed to integrate into a host cell genome by homologous recombination where expression of the gene can be affected. In one embodiment, regulatory sequences are introduced that either increase or decrease expression of an endogenous sequence. Accordingly, a 19459 protein can be produced

in a cell not normally producing it. Alternatively, increased expression of 19459 protein can be effected in a cell normally producing the protein at a specific level. Further, expression can be decreased or eliminated by introducing a specific regulatory sequence. The regulatory sequence can be heterologous to the protein sequence or can be a homologous sequence with a desired mutation that affects expression. Alternatively, the entire gene can be deleted. The regulatory sequence can be specific to the host cell or capable of functioning in more than one cell type. Still further, specific mutations can be introduced into any desired region of the gene to produce mutant 19459 proteins. Such mutations could be introduced, for example, into the specific functional regions such as the ligand-binding site.

In one embodiment, the host cell can be a fertilized oocyte or embryonic stem cell that can be used to produce a transgenic animal containing the altered 19459 gene. Alternatively, the host cell can be a stem cell or other early tissue precursor that gives rise to a specific subset of cells and can be used to produce transgenic tissues in an animal. See also Thomas *et al.*, *Cell 51*:503 (1987) for a description of homologous recombination vectors. The vector is introduced into an embryonic stem cell line (e.g., by electroporation) and cells in which the introduced gene has homologously recombined with the endogenous gene is selected (see e.g., Li, E. *et al.*, *Cell 69*:915 (1992)). The selected cells are then injected into a blastocyst of an animal (e.g., a mouse) to form aggregation chimeras (see e.g., Bradley, A. in *Teratocarcinomas and Embryonic Stem Cells: A Practical Approach*, E.J. Robertson, ed. (IRL, Oxford, 1987) pp. 113-152). A chimeric embryo can then be implanted into a suitable pseudopregnant female foster animal and the embryo brought to term. Progeny harboring the homologously recombined DNA in their germ cells can be used to breed animals in which all cells of the animal contain the homologously recombined DNA by germline transmission of the transgene. Methods for constructing homologous recombination vectors and homologous recombinant animals are described further in Bradley, A. (1991) *Current Opinions in Biotechnology* 2:823-829 and in PCT International Publication Nos. WO 90/11354; WO 91/01140; and WO 93/04169.

The genetically engineered host cells can be used to produce non-human transgenic animals. A transgenic animal is preferably a mammal, for example a rodent, such as a rat or mouse, in which one or more of the cells of the animal include a transgene. A transgene is exogenous DNA which is integrated into the genome of a cell

from which a transgenic animal develops and which remains in the genome of the mature animal in one or more cell types or tissues of the transgenic animal. These animals are useful for studying the function of a 19459 protein and identifying and evaluating modulators of the protein activity.

Other examples of transgenic animals include non-human primates, sheep, dogs, cows, goats, chickens, and amphibians.

In one embodiment, a host cell is a fertilized oocyte or an embryonic stem cell into which 19459 polynucleotide sequences have been introduced.

A transgenic animal can be produced by introducing nucleic acid into the male pronuclei of a fertilized oocyte, e.g., by microinjection, retroviral infection, and allowing the oocyte to develop in a pseudopregnant female foster animal. Any of the 19459 nucleotide sequences can be introduced as a transgene into the genome of a non-human animal, such as a mouse.

Any of the regulatory or other sequences useful in expression vectors can form part of the transgenic sequence. This includes intronic sequences and polyadenylation signals, if not already included. A tissue-specific regulatory sequence(s) can be operably linked to the transgene to direct expression of the protein to particular cells.

Methods for generating transgenic animals via embryo manipulation and microinjection, particularly animals such as mice, have become conventional in the art and are described, for example, in U.S. Patent Nos. 4,736,866 and 4,870,009, both by Leder *et al.*, U.S. Patent No. 4,873,191 by Wagner *et al.* and in Hogan, B., *Manipulating the Mouse Embryo*, (Cold Spring Harbor Laboratory Press, Cold Spring Harbor, N.Y., 1986). Similar methods are used for production of other transgenic animals. A transgenic founder animal can be identified based upon the presence of the transgene in its genome and/or expression of transgenic mRNA in tissues or cells of the animals. A transgenic founder animal can then be used to breed additional animals carrying the transgene. Moreover, transgenic animals carrying a transgene can further be bred to other transgenic animals carrying other transgenes. A transgenic animal also includes animals in which the entire animal or tissues in the animal have been produced using the homologously recombinant host cells described herein.

In another embodiment, transgenic non-human animals can be produced which contain selected systems which allow for regulated expression of the transgene. One example of such a system is the *cre/loxP* recombinase system of bacteriophage P1. For

a description of the *cre/loxP* recombinase system, see, e.g., Lakso *et al. PNAS 89*:6232-6236 (1992). Another example of a recombinase system is the FLP recombinase system of *S. cerevisiae* (O'Gorman *et al. Science 251*:1351-1355 (1991). If a *cre/loxP* recombinase system is used to regulate expression of the transgene, animals containing transgenes encoding both the *Cre* recombinase and a selected protein is required. Such animals can be provided through the construction of "double" transgenic animals, e.g., by mating two transgenic animals, one containing a transgene encoding a selected protein and the other containing a transgene encoding a recombinase.

Clones of the non-human transgenic animals described herein can also be produced according to the methods described in Wilmut, I. *et al. Nature 385*:810-813 (1997) and PCT International Publication Nos. WO 97/07668 and WO 97/07669. In brief, a cell, e.g., a somatic cell, from the transgenic animal can be isolated and induced to exit the growth cycle and enter G_o phase. The quiescent cell can then be fused, e.g., through the use of electrical pulses, to an enucleated oocyte from an animal of the same species from which the quiescent cell is isolated. The reconstructed oocyte is then cultured such that it develops to morula or blastocyst and then transferred to a pseudopregnant female foster animal. The offspring born of this female foster animal will be a clone of the animal from which the cell, e.g., the somatic cell, is isolated.

Transgenic animals containing recombinant cells that express the polypeptides described herein are useful to conduct the assays described herein in an *in vivo* context. Accordingly, the various physiological factors that are present *in vivo* and that could affect ligand binding, protein (e.g., receptor) activation, and signal transduction, may not be evident from *in vitro* cell-free or cell-based assays. Accordingly, it is useful to provide non-human transgenic animals to assay *in vivo* protein (e.g., receptor) function, including ligand interaction, the effect of specific mutant 19459 proteins on 19459 protein function and ligand interaction, and the effect of chimeric 19459 proteins. It is also possible to assess the effect of null mutations, that is mutations that substantially or completely eliminate one or more protein functions.

In general, methods for producing transgenic animals include introducing a nucleic acid sequence according to the present invention, the nucleic acid sequence capable of expressing the protein in a transgenic animal, into a cell in culture or *in vivo*. When introduced *in vivo*, the nucleic acid is introduced into an intact organism such that one or more cell types and, accordingly, one or more tissue types, express

the nucleic acid encoding the protein. Alternatively, the nucleic acid can be introduced into virtually all cells in an organism by transfecting a cell in culture, such as an embryonic stem cell, as described herein for the production of transgenic animals, and this cell can be used to produce an entire transgenic organism. As described, in a further embodiment, the host cell can be a fertilized oocyte. Such cells are then allowed to develop in a female foster animal to produce the transgenic organism.

Pharmaceutical compositions

The nucleic acid molecules, protein (particularly fragments such as the amino terminal extracellular domain), modulators of the protein, and antibodies (also referred to herein as "active compounds") can be incorporated into pharmaceutical compositions suitable for administration to a subject, e.g., a human. Such compositions typically comprise the nucleic acid molecule, protein, modulator, or antibody and a pharmaceutically acceptable carrier.

As used herein the language "pharmaceutically acceptable carrier" is intended to include any and all solvents, dispersion media, coatings, antibacterial and antifungal agents, isotonic and absorption delaying agents, and the like, compatible with pharmaceutical administration. The use of such media and agents for pharmaceutically active substances is well known in the art. Except insofar as any conventional media or agent is incompatible with the active compound, such media can be used in the compositions of the invention. Supplementary active compounds can also be incorporated into the compositions. A pharmaceutical composition of the invention is formulated to be compatible with its intended route of administration. Examples of routes of administration include parenteral, e.g., intravenous, intradermal, subcutaneous, oral (e.g., inhalation), transdermal (topical), transmucosal, and rectal administration. Solutions or suspensions used for parenteral, intradermal, or subcutaneous application can include the following components: a sterile diluent such as water for injection, saline solution, fixed oils, polyethylene glycols, glycerine, propylene glycol or other synthetic solvents; antibacterial agents such as benzyl alcohol or methyl parabens; antioxidants such as ascorbic acid or sodium bisulfite; chelating agents such as ethylenediaminetetraacetic acid; buffers such as acetates, citrates or phosphates and agents for the adjustment of tonicity such as sodium chloride or dextrose. pH can be

adjusted with acids or bases, such as hydrochloric acid or sodium hydroxide. The parenteral preparation can be enclosed in ampules, disposable syringes or multiple dose vials made of glass or plastic.

Pharmaceutical compositions suitable for injectable use include sterile aqueous solutions (where water soluble) or dispersions and sterile powders for the extemporaneous preparation of sterile injectable solutions or dispersion. For intravenous administration, suitable carriers include physiological saline, bacteriostatic water, Cremophor EL™ (BASF, Parsippany, NJ) or phosphate buffered saline (PBS). In all cases, the composition must be sterile and should be fluid to the extent that easy syringability exists. It must be stable under the conditions of manufacture and storage and must be preserved against the contaminating action of microorganisms such as bacteria and fungi. The carrier can be a solvent or dispersion medium containing, for example, water, ethanol, polyol (for example, glycerol, propylene glycol, and liquid polyethylene glycol, and the like), and suitable mixtures thereof. The proper fluidity can be maintained, for example, by the use of a coating such as lecithin, by the maintenance of the required particle size in the case of dispersion and by the use of surfactants. Prevention of the action of microorganisms can be achieved by various antibacterial and antifungal agents, for example, parabens, chlorobutanol, phenol, ascorbic acid, thimerosal, and the like. In many cases, it will be preferable to include isotonic agents, for example, sugars, polyalcohols such as mannitol, sorbitol, sodium chloride in the composition. Prolonged absorption of the injectable compositions can be brought about by including in the composition an agent which delays absorption, for example, aluminum monostearate and gelatin.

Sterile injectable solutions can be prepared by incorporating the active compound (e.g., a 19459 protein or anti-19459 antibody) in the required amount in an appropriate solvent with one or a combination of ingredients enumerated above, as required, followed by filtered sterilization. Generally, dispersions are prepared by incorporating the active compound into a sterile vehicle which contains a basic dispersion medium and the required other ingredients from those enumerated above. In the case of sterile powders for the preparation of sterile injectable solutions, the preferred methods of preparation are vacuum drying and freeze-drying which yields a powder of the active ingredient plus any additional desired ingredient from a previously sterile-filtered solution thereof.

Oral compositions generally include an inert diluent or an edible carrier. They can be enclosed in gelatin capsules or compressed into tablets. For oral administration, the agent can be contained in enteric forms to survive the stomach or further coated or mixed to be released in a particular region of the GI tract by known methods. For the purpose of oral therapeutic administration, the active compound can be incorporated with excipients and used in the form of tablets, troches, or capsules. Oral compositions can also be prepared using a fluid carrier for use as a mouthwash, wherein the compound in the fluid carrier is applied orally and swished and expectorated or swallowed. Pharmaceutically compatible binding agents, and/or adjuvant materials can be included as part of the composition. The tablets, pills, capsules, troches and the like can contain any of the following ingredients, or compounds of a similar nature: a binder such as microcrystalline cellulose, gum tragacanth or gelatin; an excipient such as starch or lactose, a disintegrating agent such as alginic acid, Primogel, or corn starch; a lubricant such as magnesium stearate or Sterotes; a glidant such as colloidal silicon dioxide; a sweetening agent such as sucrose or saccharin; or a flavoring agent such as peppermint, methyl salicylate, or orange flavoring.

For administration by inhalation, the compounds are delivered in the form of an aerosol spray from pressured container or dispenser which contains a suitable propellant, e.g., a gas such as carbon dioxide, or a nebulizer.

Systemic administration can also be by transmucosal or transdermal means. For transmucosal or transdermal administration, penetrants appropriate to the barrier to be permeated are used in the formulation. Such penetrants are generally known in the art, and include, for example, for transmucosal administration, detergents, bile salts, and fusidic acid derivatives. Transmucosal administration can be accomplished through the use of nasal sprays or suppositories. For transdermal administration, the active compounds are formulated into ointments, salves, gels, or creams as generally known in the art.

The compounds can also be prepared in the form of suppositories (e.g., with conventional suppository bases such as cocoa butter and other glycerides) or retention enemas for rectal delivery.

In one embodiment, the active compounds are prepared with carriers that will protect the compound against rapid elimination from the body, such as a controlled release formulation, including implants and microencapsulated delivery systems.

Biodegradable, biocompatible polymers can be used, such as ethylene vinyl acetate, polyanhydrides, polyglycolic acid, collagen, polyorthoesters, and polylactic acid. Methods for preparation of such formulations will be apparent to those skilled in the art. The materials can also be obtained commercially from Alza Corporation and Nova Pharmaceuticals, Inc. Liposomal suspensions (including liposomes targeted to infected cells with monoclonal antibodies to viral antigens) can also be used as pharmaceutically acceptable carriers. These can be prepared according to methods known to those skilled in the art, for example, as described in U.S. Patent No. 4,522,811.

It is especially advantageous to formulate oral or parenteral compositions in dosage unit form for ease of administration and uniformity of dosage. "Dosage unit form" as used herein refers to physically discrete units suited as unitary dosages for the subject to be treated; each unit containing a predetermined quantity of active compound calculated to produce the desired therapeutic effect in association with the required pharmaceutical carrier. The specification for the dosage unit forms of the invention are dictated by and directly dependent on the unique characteristics of the active compound and the particular therapeutic effect to be achieved, and the limitations inherent in the art of compounding such an active compound for the treatment of individuals.

The nucleic acid molecules of the invention can be inserted into vectors and used as gene therapy vectors. Gene therapy vectors can be delivered to a subject by, for example, intravenous injection, local administration (U.S. 5,328,470) or by stereotactic injection (see e.g., Chen *et al.*, *PNAS 91*:3054-3057 (1994)). The pharmaceutical preparation of the gene therapy vector can include the gene therapy vector in an acceptable diluent, or can comprise a slow release matrix in which the gene delivery vehicle is imbedded. Alternatively, where the complete gene delivery vector can be produced intact from recombinant cells, e.g. retroviral vectors, the pharmaceutical preparation can include one or more cells which produce the gene delivery system.

As defined herein, a therapeutically effective amount of protein or polypeptide (i.e., an effective dosage) ranges from about 0.001 to 30 mg/kg body weight, preferably about 0.01 to 25 mg/kg body weight, more preferably about 0.1 to 20 mg/kg body weight, and even more preferably about 1 to 10 mg/kg, 2 to 9 mg/kg, 3 to 8 mg/kg, 4 to 7 mg/kg, or 5 to 6 mg/kg body weight.

The skilled artisan will appreciate that certain factors may influence the dosage required to effectively treat a subject, including but not limited to the severity

of the disease or disorder, previous treatments, the general health and/or age of the subject, and other diseases present. Moreover, treatment of a subject with a therapeutically effective amount of a protein, polypeptide, or antibody can include a single treatment or, preferably, can include a series of treatments. In a preferred example, a subject is treated with antibody, protein, or polypeptide in the range of between about 0.1 to 20 mg/kg body weight, one time per week for between about 1 to 10 weeks, preferably between 2 to 8 weeks, more preferably between about 3 to 7 weeks, and even more preferably for about 4, 5, or 6 weeks. It will also be appreciated that the effective dosage of antibody, protein, or polypeptide used for treatment may increase or decrease over the course of a particular treatment. Changes in dosage may result and become apparent from the results of diagnostic assays as described herein.

The present invention encompasses agents which modulate expression or activity. An agent may, for example, be a small molecule. For example, such small molecules include, but are not limited to, peptides, peptidomimetics, amino acids, amino acid analogs, polynucleotides, polynucleotide analogs, nucleotides, nucleotide analogs, organic or inorganic compounds (i.e., including heteroorganic and organometallic compounds) having a molecular weight less than about 10,000 grams per mole, organic or inorganic compounds having a molecular weight less than about 5,000 grams per mole, organic or inorganic compounds having a molecular weight less than about 1,000 grams per mole, organic or inorganic compounds having a molecular weight less than about 500 grams per mole, and salts, esters, and other pharmaceutically acceptable forms of such compounds.

It is understood that appropriate doses of small molecule agents depends upon a number of factors within the ken of the ordinarily skilled physician, veterinarian, or researcher. The dose(s) of the small molecule will vary, for example, depending upon the identity, size, and condition of the subject or sample being treated, further depending upon the route by which the composition is to be administered, if applicable, and the effect which the practitioner desires the small molecule to have upon the nucleic acid or polypeptide of the invention. Exemplary doses include milligram or microgram amounts of the small molecule per kilogram of subject or sample weight (e.g., about 1 microgram per kilogram to about 500 milligrams per kilogram, about 100 micrograms per kilogram to about 5 milligrams per kilogram, or

about 1 microgram per kilogram to about 50 micrograms per kilogram. It is furthermore understood that appropriate doses of a small molecule depend upon the potency of the small molecule with respect to the expression or activity to be modulated. Such appropriate doses may be determined using the assays described herein. When one or more of these small molecules is to be administered to an animal (e.g., a human) in order to modulate expression or activity of a polypeptide or nucleic acid of the invention, a physician, veterinarian, or researcher may, for example, prescribe a relatively low dose at first, subsequently increasing the dose until an appropriate response is obtained. In addition, it is understood that the specific dose level for any particular animal subject will depend upon a variety of factors including the activity of the specific compound employed, the age, body weight, general health, gender, and diet of the subject, the time of administration, the route of administration, the rate of excretion, any drug combination, and the degree of expression or activity to be modulated.

The pharmaceutical compositions can be included in a container, pack, or dispenser together with instructions for administration.

This invention may be embodied in many different forms and should not be construed as limited to the embodiments set forth herein; rather, these embodiments are provided so that this disclosure will fully convey the invention to those skilled in the art. Many modifications and other embodiments of the invention will come to mind in one skilled in the art to which this invention pertains having the benefit of the teachings presented in the foregoing description. Although specific terms are employed, they are used as in the art unless otherwise indicated.

Other Embodiments

In another aspect, the invention features, a method of analyzing a plurality of capture probes. The method can be used, e.g., to analyze gene expression. The method includes: providing a two dimensional array having a plurality of addresses, each address of the plurality being positionally distinguishable from each other address of the plurality, and each address of the plurality having a unique capture probe, e.g., a nucleic acid or peptide sequence; contacting the array with a 19459, preferably purified, nucleic acid, preferably purified, polypeptide, preferably purified, or antibody, and thereby evaluating the plurality of capture probes. Binding, e.g., in the case of a nucleic acid, hybridization with a capture probe at an address of the plurality, is detected, e.g., by

signal generated from a label attached to the 19459 nucleic acid, polypeptide, or antibody.

The capture probes can be a set of nucleic acids from a selected sample, e.g., a sample of nucleic acids derived from a control or non-stimulated tissue or cell. The method can include contacting the 19459 nucleic acid, polypeptide, or antibody with a first array having a plurality of capture probes and a second array having a different plurality of capture probes. The results of each hybridization can be compared, e.g., to analyze differences in expression between a first and second sample. The first plurality of capture probes can be from a control sample, e.g., a wild type, normal, or non-diseased, non-stimulated, sample, e.g., a biological fluid, tissue, or cell sample. The second plurality of capture probes can be from an experimental sample, e.g., a mutant type, at risk, disease-state or disorder-state, or stimulated, sample, e.g., a biological fluid, tissue, or cell sample.

The plurality of capture probes can be a plurality of nucleic acid probes each of which specifically hybridizes, with an allele of 19459. Such methods can be used to diagnose a subject, e.g., to evaluate risk for a disease or disorder, to evaluate suitability of a selected treatment for a subject, to evaluate whether a subject has a disease or disorder. 19459 is associated with 7 transmembrane protein activity, thus it is useful for disorders associated with abnormal 7 transmembrane protein signaling.

The method can be used to detect SNPs.

In another aspect, the invention features, a method of analyzing a plurality of probes. The method is useful, e.g., for analyzing gene expression. The method includes: providing a two dimensional array having a plurality of addresses, each address of the plurality being positionally distinguishable from each other address of the plurality having a unique capture probe, e.g., wherein the capture probes are from a cell or subject which express or misexpress 19459 or from a cell or subject in which a 19459 mediated response has been elicited, e.g., by contact of the cell with 19459 nucleic acid or protein, or administration to the cell or subject 19459 nucleic acid or protein; contacting the array with one or more inquiry probe, wherein an inquiry probe can be a nucleic acid, polypeptide, or antibody (which is preferably other than 19459 nucleic acid, polypeptide, or antibody); providing a two dimensional array having a plurality of addresses, each address of the plurality being positionally distinguishable from each other address of the plurality, and each address of the plurality having a unique capture probe, e.g., wherein

the capture probes are from a cell or subject which does not express 19459 (or does not express as highly as in the case of the 19459 positive plurality of capture probes) or from a cell or subject which in which a 19459 mediated response has not been elicited (or has been elicited to a lesser extent than in the first sample); contacting the array with one or more inquiry probes (which is preferably other than a 19459 nucleic acid, polypeptide, or antibody), and thereby evaluating the plurality of capture probes. Binding, e.g., in the case of a nucleic acid, hybridization with a capture probe at an address of the plurality, is detected, e.g., by signal generated from a label attached to the nucleic acid, polypeptide, or antibody.

In another aspect, the invention features, a method of analyzing 19459, e.g., analyzing structure, function, or relatedness to other nucleic acid or amino acid sequences. The method includes: providing a 19459 nucleic acid or amino acid sequence; comparing the 19459 sequence with one or more preferably a plurality of sequences from a collection of sequences, e.g., a nucleic acid or protein sequence database; to thereby analyze 19459.

Preferred databases include GenBank™. The method can include evaluating the sequence identity between a 19459 sequence and a database sequence. The method can be performed by accessing the database at a second site, e.g., over the internet.

In another aspect, the invention features, a set of oligonucleotides, useful, e.g., for identifying SNP's, or identifying specific alleles of 19459. The set includes a plurality of oligonucleotides, each of which has a different nucleotide at an interrogation position, e.g., an SNP or the site of a mutation. In a preferred embodiment, the oligonucleotides of the plurality are identical in sequence with one another (except for differences in length). The oligonucleotides can be provided with different labels, such that an oligonucleotide that hybridizes to one allele provides a signal that is distinguishable from an oligonucleotide which hybridizes to a second allele.

This invention is further illustrated by the following examples which should not be construed as limiting. The contents of all references, patents and published patent applications cited throughout this application are incorporated herein by reference.

EXAMPLES

Example 1: Identification and Characterization of Human 19459 cDNAs

The human 19459 sequence (Figure 1; SEQ ID NO:1), which is approximately 1737 nucleotides long including untranslated regions, contains a predicted methionine-initiated coding sequence of about 1191 nucleotides (nucleotides 527-1717 of SEQ ID NO:1; SEQ ID NO:3). The coding sequence encodes a 397 amino acid protein (shown in SEQ ID NO:2).

Example 2: Tissue Distribution of 19459 mRNA

Expression levels of 19459 in various tissue and cell types were determined by quantitative RT (reverse transcriptase PCR (Taqman® brand PCR kit, Applied Biosystems). The quantitative RT-PCR reactions were performed according to the kit manufacturer's instructions. The results are shown in figures 5-7. 19459 expression was detectable in several tissue and cell types including but not limited to the following: heart, kidney, muscle, liver (including fibrotic liver), placenta, lung, tonsil, brain, colon, spleen, stellate cells, lung fibroblasts, Th1, Th2, and Th3 cells, granulocytes, CD19 positive cells, peripheral blood mononuclear cells, bronchial epithelial cells, CD 34 positive cells (from mobilized peripheral blood, adult bone marrow, and mobilized bone marrow), erythroid cells, megakaryocytes, neutrophils, osteoclasts, osteoblasts, thyroid, and testis. Quantitative PCR also demonstrated that 19459 was expressed in normal vein and human umbilical vein endothelial cells (maintained under static or shear force conditions).

Northern blot hybridizations with various RNA samples are performed under standard conditions and washed under stringent conditions, i.e., 0.2 X SSC at 65°C. A DNA probe corresponding to all or a portion of the 19459 cDNA (SEQ ID NO:1) can be used. The DNA is radioactively labeled with 32P-dCTP using the Prime-It Kit (Stratagene, La Jolla, CA) according to the instructions of the supplier. Filters containing mRNA from mouse hematopoietic and endocrine tissues, and cancer cell lines (Clontech, Palo Alto, CA) are probed in ExpressHyb hybridization solution (Clontech) and washed at high stringency according to manufacturer's recommendations.

Add a description of any expression panel results. If the results are generated using TaqMan, add the following:

Example 3: Recombinant Expression of 19459 in Bacterial Cells

In this example, 19459 is expressed as a recombinant glutathione-S-transferase (GST) fusion polypeptide in E. coli and the fusion polypeptide is isolated and characterized. Specifically, 19459 is fused to GST and this fusion polypeptide is expressed in E. coli, e.g., strain PEB199. Expression of the GST-19459 fusion protein in PEB199 is induced with IPTG. The recombinant fusion polypeptide is purified from crude bacterial lysates of the induced PEB199 strain by affinity chromatography on glutathione beads. Using polyacrylamide gel electrophoretic analysis of the polypeptide purified from the bacterial lysates, the molecular weight of the resultant fusion polypeptide is determined.

Example 4: Expression of Recombinant 19459 Protein in COS Cells

To express the 19459 gene in COS cells, the pcDNA/Amp vector by Invitrogen Corporation (San Diego, CA) is used. This vector contains an SV40 origin of replication, an ampicillin resistance gene, an E. coli replication origin, a CMV promoter followed by a polylinker region, and an SV40 intron and polyadenylation site. A DNA fragment encoding the entire 19459 protein and an HA tag (Wilson et al. (1984) Cell 37:767) or a FLAG tag fused in-frame to its 3' end of the fragment is cloned into the polylinker region of the vector, thereby placing the expression of the recombinant protein under the control of the CMV promoter.

To construct the plasmid, the 19459 DNA sequence is amplified by PCR using two primers. The 5' primer contains the restriction site of interest followed by approximately twenty nucleotides of the 19459 coding sequence starting from the initiation codon; the 3' end sequence contains complementary sequences to the other restriction site of interest, a translation stop codon, the HA tag or FLAG tag and the last 20 nucleotides of the 19459 coding sequence. The PCR amplified fragment and the pCDNA/Amp vector are digested with the appropriate restriction enzymes and the vector is dephosphorylated using the CIAP enzyme (New England Biolabs, Beverly, MA). Preferably the two restriction sites chosen are different so that the 19459 gene is inserted in the correct orientation. The ligation mixture is transformed into E. coli cells (strains HB101, DH5α, SURE, available from Stratagene Cloning Systems, La Jolla, CA, can

be used), the transformed culture is plated on ampicillin media plates, and resistant colonies are selected. Plasmid DNA is isolated from transformants and examined by restriction analysis for the presence of the correct fragment.

COS cells are subsequently transfected with the 19459-pcDNA/Amp plasmid DNA using the calcium phosphate or calcium chloride co-precipitation methods, DEAE-dextran-mediated transfection, lipofection, or electroporation. Other suitable methods for transfecting host cells can be found in Sambrook, J., Fritsh, E. F., and Maniatis, T. Molecular Cloning: A Laboratory Manual. 2nd, ed., Cold Spring Harbor Laboratory, Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY, 1989. The expression of the 19459 polypeptide is detected by radiolabelling (35S-methionine or 35S-cysteine available from NEN, Boston, MA, can be used) and immunoprecipitation (Harlow, E. and Lane, D. Antibodies: A Laboratory Manual, Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY, 1988) using an HA specific monoclonal antibody. Briefly, the cells are labeled for 8 hours with 35S-methionine (or 35S-cysteine). The culture media are then collected and the cells are lysed using detergents (RIPA buffer, 150 mM NaCl, 1% NP-40, 0.1% SDS, 0.5% DOC, 50 mM Tris, pH 7.5). Both the cell lysate and the culture media are precipitated with an HA specific monoclonal antibody. Precipitated polypeptides are then analyzed by SDS-PAGE.

Alternatively, DNA containing the 19459 coding sequence is cloned directly into the polylinker of the pCDNA/Amp vector using the appropriate restriction sites. The resulting plasmid is transfected into COS cells in the manner described above, and the expression of the 19459 polypeptide is detected by radiolabelling and immunoprecipitation using a 19459 specific monoclonal antibody.

Applicant's or agent's file reference	35800/209004	International application No.	PCT/US01/

INDICATIONS RELATING TO DEPOSITED MICROORGANISM OR OTHER BIOLOGICAL MATERIAL

(PCT Rule 13*bis*)

A. The indications made below relate to the deposited microorganism or other biological material referred to in the description on page 4, line 2

B. **IDENTIFICATION OF DEPOSIT** Further deposits are identified on an additional sheet ☐

Name of depository institution
American Type Culture Collection

Address of depositary institution *(including postal code and country)*
10801 University Blvd.
Manassas, VA 20110-2209 USA

Date of deposit	Accession Number
	PTA-

C. **ADDITIONAL INDICATIONS** *(leave blank if not applicable)* This information is continued on an additional sheet ☐

Page 9, line 24; page 11, line 17; page 88, lines 9, 14, 18, 22, 25 and 29; page 89, lines 8 and 12; page 90, lines 2, 6, 11, 28 and 31; page 91, lines 2 and 6.

D. **DESIGNATED STATES FOR WHICH INDICATIONS ARE MADE** *(if the indicators are not for all designated States)*

E. **SEPARATE FURNISHING OF INDICATIONS** *(leave blank if not applicable)*

The indications listed below will be submitted to the International Bureau later *(specify the general nature of the indications e.g., "Accession Number of Deposit")*

Date of Deposit and Accession Number of Deposit

For receiving Office use only	For International Bureau use only
☐ This sheet was received with the international application	☐ This sheet was received with the International Bureau on:
Authorized officer	Authorized officer

THAT WHICH IS CLAIMED:

1. An isolated nucleic acid molecule selected from the group consisting of:

a) a nucleic acid molecule comprising a nucleotide sequence which is at least 60% identical to the nucleotide sequence of SEQ ID NO:1, SEQ ID NO:3, or the nucleotide sequence of the cDNA insert of the plasmid deposited with ATCC as Patent Deposit Number _____, wherein said nucleotide sequence encodes a polypeptide having biological activity;

b) a nucleic acid molecule comprising a fragment of at least 20 nucleotides of the nucleotide sequence of SEQ ID NO:1, SEQ ID NO:3, or the nucleotide sequence of the cDNA insert of the plasmid deposited with ATCC as Patent Deposit Number _____;

c) a nucleic acid molecule which encodes a polypeptide comprising the amino acid sequence of SEQ ID NO:2, or the amino acid sequence encoded by the cDNA insert of the plasmid deposited with the ATCC as Patent Deposit Number _____;

d) a nucleic acid molecule which encodes a fragment of a polypeptide comprising the amino acid sequence of SEQ ID NO:2, or the amino acid sequence encoded by the cDNA insert of the plasmid deposited with the ATCC as Patent Deposit Number _____, wherein the fragment comprises at least 15 contiguous amino acids of SEQ ID NO:2, or the amino acid sequence encoded by the cDNA insert of the plasmid deposited with the ATCC as Patent Deposit Number _____;

e) a nucleic acid molecule encoding a naturally occurring allelic variant of a biologically active polypeptide comprising the amino acid sequence of SEQ ID NO:2, or the amino acid sequence encoded by the cDNA insert of the plasmid deposited with the ATCC as Patent Deposit Number _____, wherein the nucleic acid molecule hybridizes to a nucleic acid molecule comprising the complement of SEQ ID NO:1 or SEQ ID NO:3 under stringent conditions;

f) a nucleic acid molecule encoding the amino sequence shown as amino acids 44-397, of SEQ ID NO:2; and

g) a nucleic acid molecule comprising the complement of a), b), c), d), e), or f).

2. The isolated nucleic acid molecule of claim 1, which is selected from the group consisting of:

a) a nucleic acid comprising the nucleotide sequence of SEQ ID NO:1, SEQ ID NO:3, the nucleotide sequence of the cDNA insert of the plasmid deposited with ATCC as Patent Deposit Number _____, or a complement thereof;

b) a nucleic acid molecule encoding a polypeptide comprising the amino acid sequence of SEQ ID NO:2, or the amino acid sequence encoded by the cDNA insert of the plasmid deposited with the ATCC as Patent Deposit Number _____, or a complement thereof; and

c) a nucleic acid molecule comprising a nucleotide sequence consisting of nucleotides 1-508, 1161-1305, or 1491-1737 of SEQ ID NO:1.

3. The nucleic acid molecule of claim 1 further comprising vector nucleic acid sequences.

4. The nucleic acid molecule of claim 1 further comprising nucleic acid sequences encoding a heterologous polypeptide.

5. A host cell which contains the nucleic acid molecule of claim 1.

6. The host cell of claim 5 which is a mammalian host cell.

7. A non-human mammalian host cell containing the nucleic acid molecule of claim 1.

8. An isolated polypeptide selected from the group consisting of:

a) a biologically active polypeptide which is encoded by a nucleic acid molecule comprising a nucleotide sequence which is at least 60% identical to a nucleic acid comprising the nucleotide sequence of SEQ ID NO:1, SEQ ID NO:3, or

the nucleotide sequence of the cDNA insert of the plasmid deposited with ATCC as Patent Deposit Number _____;

b) a naturally occurring allelic variant of a polypeptide comprising the amino acid sequence of SEQ ID NO:2, or the amino acid sequence encoded by the cDNA insert of the plasmid deposited with the ATCC as Patent Deposit Number _____, wherein the polypeptide is encoded by a nucleic acid molecule which hybridizes to a nucleic acid molecule comprising the complement of SEQ ID NO:1 or SEQ ID NO:3 under stringent conditions;

c) a fragment of a polypeptide comprising the amino acid sequence of SEQ ID NO:2, or the amino acid sequence encoded by the cDNA insert of the plasmid deposited with the ATCC as Patent Deposit Number _____, wherein the fragment comprises at least 15 contiguous amino acids of SEQ ID NO:2; and

d) a polypeptide having at least 60% sequence identity to the amino acid sequence SEQ ID NO:2, wherein the polypeptide has biological activity.

9. The isolated polypeptide of claim 8 comprising the amino acid sequence of SEQ ID NO:2.

10. The polypeptide of claim 8 further comprising heterologous amino acid sequences.

11. An antibody which selectively binds to a polypeptide of claim 8.

12. A method for producing a polypeptide selected from the group consisting of:

a) a polypeptide comprising the amino acid sequence of SEQ ID NO:2, or the amino acid sequence encoded by the cDNA insert of the plasmid deposited with the ATCC as Patent Deposit Number _____;

b) a polypeptide comprising a fragment of the amino acid sequence of SEQ ID NO:2, or the amino acid sequence encoded by the cDNA insert of the plasmid deposited with the ATCC as Patent Deposit Number _____, wherein the fragment comprises at least 15 contiguous amino acids of SEQ ID NO:2, or the

amino acid sequence encoded by the cDNA insert of the plasmid deposited with the ATCC as Patent Deposit Number _____;

c) a biologically active naturally occurring allelic variant of a polypeptide comprising the amino acid sequence of SEQ ID NO:2, or the amino acid sequence encoded by the cDNA insert of the plasmid deposited with the ATCC as Patent Deposit Number _____, wherein the polypeptide is encoded by a nucleic acid molecule which hybridizes to a nucleic acid molecule comprising the complement of SEQ ID NO:1 or SEQ ID NO:3;

d) a polypeptide having at least 60% sequence identity to the amino acid sequence of SEQ ID NO:2, wherein said polypeptide has biological activity;

comprising culturing the host cell of claim 5 under conditions in which the nucleic acid molecule is expressed.

13. A method for detecting the presence of a polypeptide of claim 8 in a sample, comprising:

a) contacting the sample with a compound which selectively binds to a polypeptide of claim 8; and

b) determining whether the compound binds to the polypeptide in the sample.

14. The method of claim 13, wherein the compound which binds to the polypeptide is an antibody.

15. A kit comprising a compound which selectively binds to a polypeptide of claim 8 and instructions for use.

16. A method for detecting the presence of a nucleic acid molecule of claim 1 in a sample, comprising the steps of:

a) contacting the sample with a nucleic acid probe or primer which selectively hybridizes to the nucleic acid molecule; and

b) determining whether the nucleic acid probe or primer binds to a nucleic acid molecule in the sample.

17. The method of claim 16, wherein the sample comprises mRNA molecules and is contacted with a nucleic acid probe.

18. A kit comprising a compound which selectively hybridizes to a nucleic acid molecule of claim 1 and instructions for use.

19. A method for identifying a compound which binds to a polypeptide of claim 8 comprising the steps of:

a) contacting a polypeptide, or a cell expressing a polypeptide of claim 8 with a test compound; and

b) determining whether the polypeptide binds to the test compound.

20. The method of claim 19, wherein the binding of the test compound to the polypeptide is detected by a method selected from the group consisting of:

a) detection of binding by direct detecting of test compound/polypeptide binding;

b) detection of binding using a competition binding assay;

c) detection of binding using an assay for 39406-mediated signaling.

21. A method for modulating the activity of a polypeptide of claim 8 comprising contacting a polypeptide or a cell expressing a polypeptide of claim 8 with a compound which binds to the polypeptide in a sufficient concentration to modulate the activity of the polypeptide.

22. A method for identifying a compound which modulates the activity of a polypeptide of claim 8, comprising:

a) contacting a polypeptide of claim 8 with a test compound; and

b) determining the effect of the test compound on the activity of the polypeptide to thereby identify a compound that modulates the activity of the polypeptide.

23. The method of claim 16, wherein said sample is from heart, liver, or kidney.

24. The method of claim 21, wherein said activity is modulated in heart, liver, or kidney.

FIG. 1A.

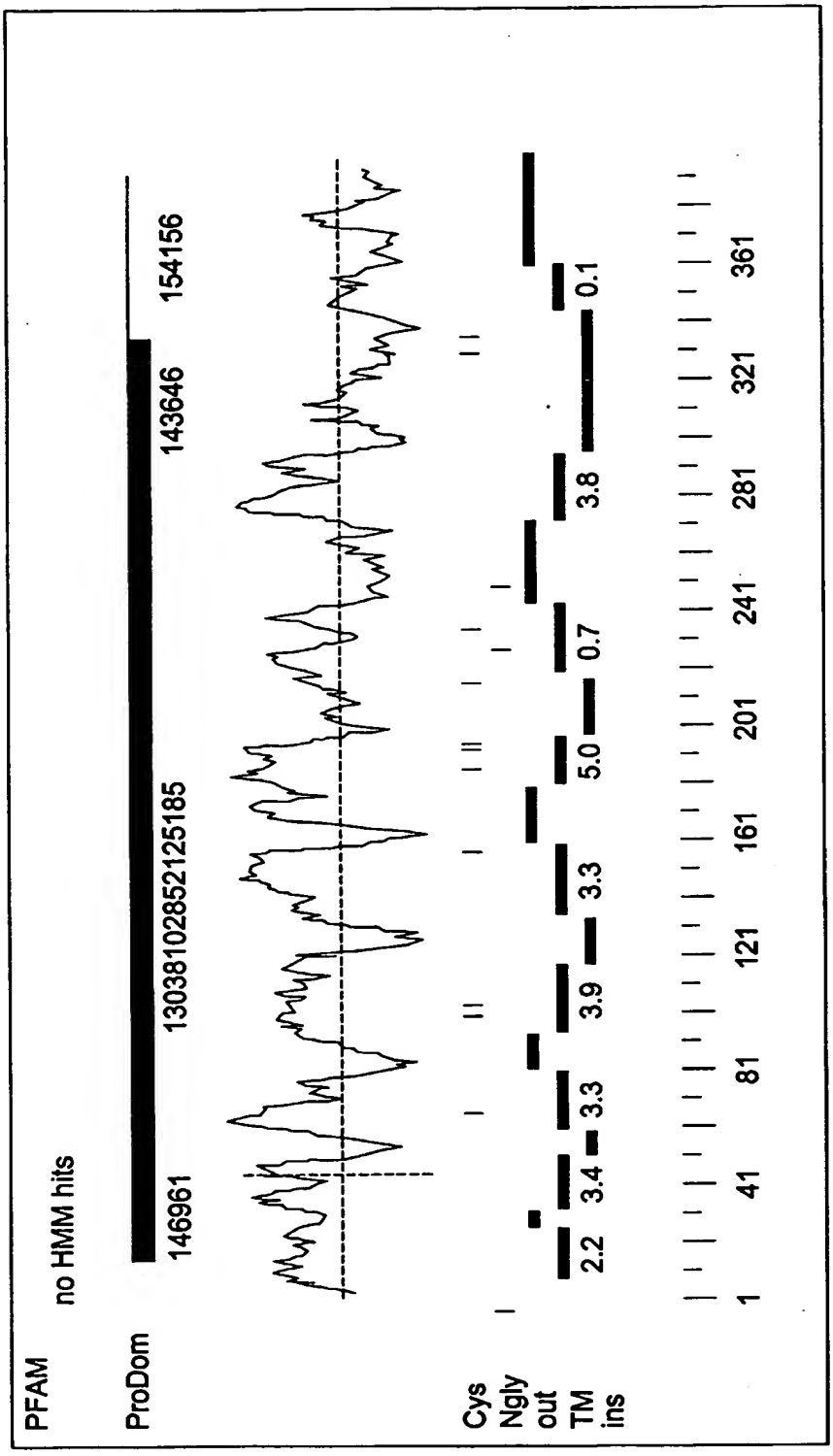
Input file 19459cons; Output file 19459tra
Sequence length 1737

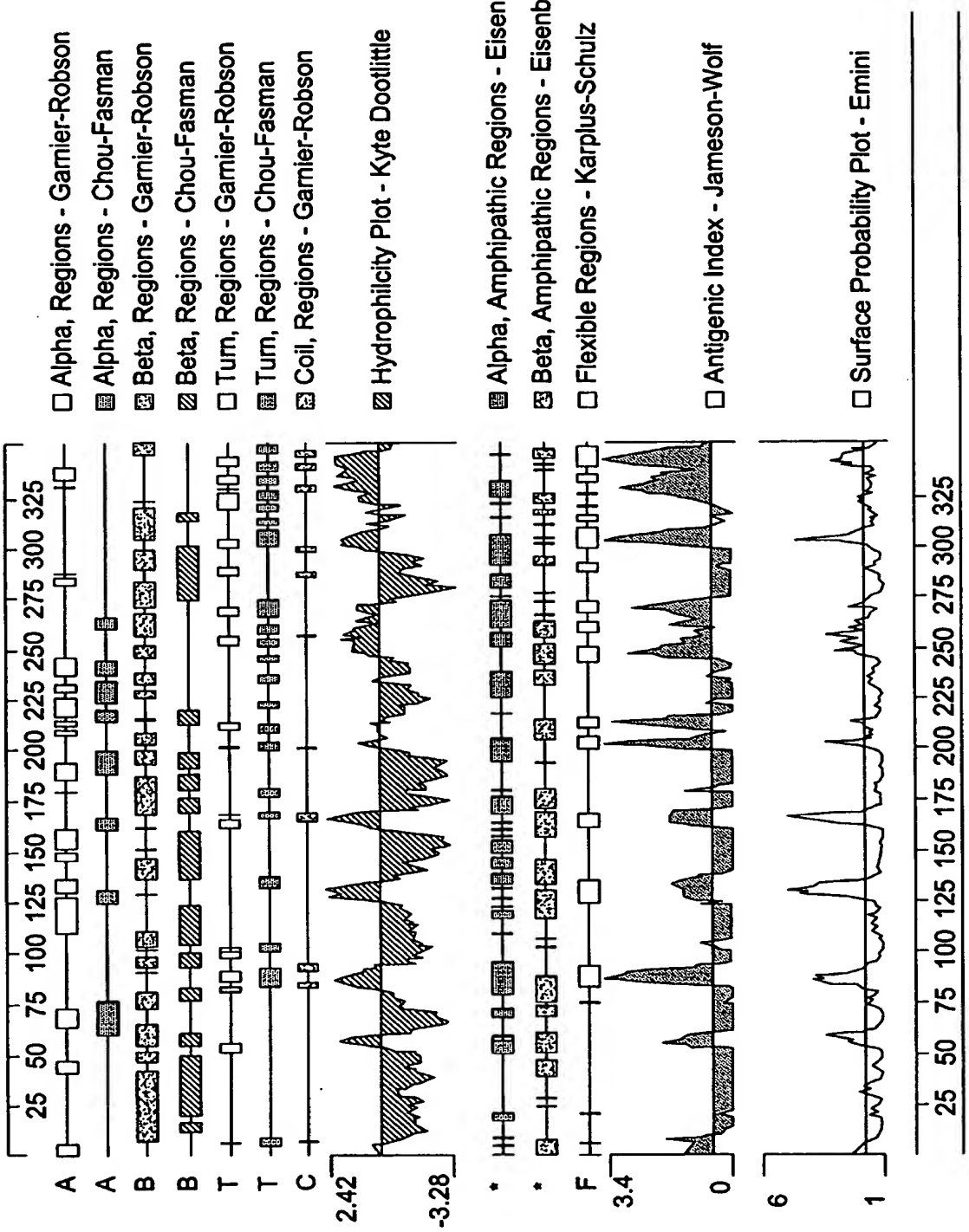
```
          10        20        30        40        50        60        70
TTCGGCACGAGGCACACCTTGGGCGTGTGTGGGAGCAGGAGAGGGATTCTCTGGGGTCCCCATTTGGAACCTGGAGACC
80        90        100       110       120       130       140       150
CACTTCTGCCAACAAGTCCTTCAGCCTCTCTGGGCCTTAGTTTCCCCCCAGTGAAGTGGGGATAGTGACAGCTCCCCCG
160       170       180       190       200       210       220       230
GGAGCCAGTGAAAACTTTGTGAAGGGAGTGATTACTGCGGGCGAAGGCGCTTGGCACAGCCACAACTGAAGCTCAGCCG
 240       250       260       270       280       290       300       310
TCGGATGTGATTATTGTTGCTATTGTTACCGCTCCGCCTGGACTGGGTGCCCGGTCGGAAAGTCCCGCTCCCCGCTCCT
 320       330       340       350       360       370       380       390
CCTCCCAGGCGCGGAAGCTTCGCGCCGCCCGCCTCGGGGGACACCTCCCGCGCCATCTCGGGGGCACTGGAGATAAGGA
 400       410       420       430       440       450       460       470
AAGGAACCTCCTCCCTGGCTGGGATGGCTGCCCGGGCGCGGGAGGAGGGCCGGGGACCCGGTGAGGGAGGACCGTAGGG
 480       490       500       510       520
                                          M   E   S   N   L   S          6
CCCGTGGGGGCTTCACGCCTGGCAGGCGGCCGGCCAGGCCTCCCTCCCTGAC ATG GAG AGT AAC CTG TCT     18   544
                                                       527
 G   L   V   P   A   A   G   L   V   P   A   L   P   P   A   V   T   L   G   L    26
GGC CTG GTC CCT GCT GCC GGG CTG GTG CCT GCG CTG CCA CCT GCT GAG ACC CTG GGG CTG    78   604
 T   A   A   Y   T   T   L   Y   A   L   L   F   F   S   V   Y   A   Q   L   V    46
ACA GCT GCC TAC ACC ACC CTG TAT GCC CTG CTC TTC TTC TCC GTC TAT GCC CAG CTG TGG   138   664
 L   V   L   L   Y   G   H   K   R   L   S   Y   Q   T   V   F   L   A   L   C    66
CTG GTG CTT CTG TAT GGG CAC AAG CGT CTC AGC TAT CAG ACG GTG TTC CTG GCC CTC TGT   198   724
 L   L   W   A   A   L   R   T   T   L   F   S   F   Y   F   R   D   T   P   R    86
CTG CTC TGG GCC GCC TTG CGT ACC ACC CTC TTC TCC TTC TAC TTC CGA GAT ACT CCC CGC   258   784
 A   N   R   L   G   P   L   P   F   W   L   L   Y   C   C   P   V   C   L   Q   106
GCC AAC CGC CTG GGG CCC TTG CCC TTC TGG CTT CTC TAC TGC TGC CCC GTC TGC CTG CAG   318   844
 F   F   T   L   T   L   M   N   L   Y   F   A   Q   V   V   F   K   A   K   V   126
TTC TTC ACC TTG ACG CTT ATG AAC CTC TAC TTT GCC CAG GTG GTG TTC AAG GCC AAG GTG   378   904
 K   R   R   P   E   M   S   R   G   L   L   A   V   R   G   A   F   V   G   A   146
AAG CGT CGG CCG GAG ATG AGC CGA GGC TTG CTC GCT GTC CGA GGG GCC TTT GTG GGG GCC   438   964
 S   L   L   F   L   L   V   N   V   L   C   A   V   L   S   H   R   R   R   A   166
TCG CTG CTC TTT CTG CTG GTG AAC GTG CTG TGT GCT GTG CTC TCC CAT CGG CGC CGG GCA   498  1024
 Q   P   W   A   L   L   L   V   R   V   L   V   S   D   S   L   F   V   I   C   186
CAG CCC TGG GCC CTG CTG CTT GTC CGC GTC CTG GTG AGC GAC TCC CTG TTC GTG ATC TGC   558  1084
 A   L   S   L   A   A   C   L   C   L   V   A   R   R   A   P   S   T   S   I   206
GCG CTG TCT CTT GCT GCC TGC CTC TGC CTC GTC GCC AGG CGG GCG CCC TCC ACT AGC ATC   618  1144
 Y   L   E   A   K   G   T   S   V   C   Q   A   A   A   M   G   G   A   M   V   226
TAC CTG GAG GCC AAG GGG ACC AGT GTG TGC CAG GCG GCC GCG ATG GGT GGC GCC ATG GTC   678  1204
 L   L   Y   A   S   R   A   C   Y   N   L   T   A   L   A   L   A   P   Q   S   246
CTG CTC TAT GCC AGC CGG GCC TGC TAC AAC CTG ACA GCA CTG GCC TTG GCC CCC CAG AGC   738  1264
 R   L   D   T   F   D   Y   D   V   Y   N   V   S   D   Q   A   D   L   V   N   266
CGG CTG GAC ACC TTC GAT TAC GAC TGG TAC AAT GTG TCT GAC CAG GCG GAC CTG GTG AAT   798  1324
 D   L   G   N   K   G   Y   L   V   F   G   L   I   L   F   V   V   E   L   L   286
GAC CTG GGG AAC AAA GGC TAC CTG GTA TTT GGC CTC ATC CTC TTC GTG TGG GAG CTA CTG   858  1384
```

```
 P  T  T  L  L  V  G  F  F  R  V  H  R  P  P  Q  D  L  S  T    306
CCC ACC ACC CTG CTG GTG GGC TTC TTC CGG GTG CAC CGG CCC CCA CAG GAC CTG AGC ACC   918  1444
 S  H  I  L  N  G  Q  V  F  A  S  R  S  Y  F  F  D  R  A  G    326
AGC CAC ATC CTC AAT GGG CAG GTC TTT GCC TCT CGG TCC TAC TTC TTT GAC CGG GCT GGG   978  1504
 H  C  E  D  E  G  C  S  V  E  H  S  R  G  E  S  T  S  N  S    346
CAC TGT GAA GAT GAG GGC TGC TCC TGG GAG CAC AGC CGG GGT GAG AGC ACC AGT ATG TCG  1038  1564
 G  S  L  G  S  G  S  W  Y  G  A  I  G  R  E  F  G  W  Y  G    366
GGC AGT CTA GGC TCT GGG AGC TGG TAT GGT GCC ATC GGG CGT GAG CCG GGC TGG TAT GGG  1098  1624
 G  S  Q  T  K  T  T  P  L  L  F  S  Q  V  P  G  P  G  G  H    386
GAC CTG GAG GCC AAG GGG ACC AGT GTG TGC CAG GCG GCC GCG ATG GGT GGC GCC ATG GTC  1158  1684
 H  H  S  L  Y  S  T  P  Q  T  *                               397
CAC CAC AGT CTC TAC TCC ACC CCA CAG ACG TGA                     1191
                                    \ 1714
TCCCCCCTGGTGCCGAATTC
```

FIG. 1B.

FIG. 2.


Analysis of No Label (397 aa)
PFAM
no HMM hits
ProDom
146961
130381028
52125185
143646
154156
Cys
Ngly
out
TM
ins
2.2
3.4
3.3
3.9
3.3
5.0
0.7
3.8
0.1
1
41
81
121
161
201
241
281
321
361


Alpha, Regions - Garnier-Robson
Alpha, Regions - Chou-Fasman
Beta, Regions - Garnier-Robson
Beta, Regions - Chou-Fasman
Turn, Regions - Garnier-Robson
Turn, Regions - Chou-Fasman
Coil, Regions - Garnier-Robson
Hydrophilicity Plot - Kyte Doolittle
Alpha, Amphipathic Regions - Eisenberg
Beta, Amphipathic Regions - Eisenberg
Flexible Regions - Karplus-Schulz
Antigenic Index - Jameson-Wolf
Surface Probability Plot - Emini
25 50 75 100 125 150 175 200 225 250 275 300 325
A
B
T
C
*
F
2.42
-3.28
3.4
0
6
1

Prosite version: Release 12.2 of February 1995

>PS00001|PDOC00001|ASN_GLYCOSYLATION N-glycosylation site.

Query: 4 NLSG 7

Query: 236 NLTA 239

Query: 257 NVSD 260

>PS00004|PDOC00004|CAMP_PHOSPHO_SITE cAMP- and cGMP-dependent protein kinase phosphorylation

Query: 54 KRLS 57

>PS00005|PDOC00005|PKC_PHOSPHO_SITE Protein kinase C phosphorylation site.

Query: 84 TPR 86

Query: 161 SHR 163

>PS00006|PDOC00006|CK2_PHOSPHO_SITE Casein kinase II phosphorylation site.

Query: 246 SRLD 249

Query: 338 SRGE 341

>PS00008|PDOC00008|MYRISTYL N-myristoylation site.

Query: 7 GLVPAA 12

Query: 13 GLVPAUL 18

Query: 25 GLTAAY 30

Query: 141 GAFVGA 146

Query: 212 GTSVCQ 217

Query: 312 GQVFAS 317

Query: 347 GSLGSG 352

Query: 366 GGSQTK 371

>PS00013|PDOC00013|PROKAR_LIPOPROTEIN Prokaryotic membraine lipoprotein lipid attachement site

Non-eukaryotic pattern
RU Additional rules:
RU (1) The cysteine must be between position 15 and 35 of the sequence in
RU consideraton
RU (2) There must be at least one charged residue (Lys or Arg) in the first
RU seven residues of the sequence

Query: 183 FVICALSLAAC 193

FIG. 4.

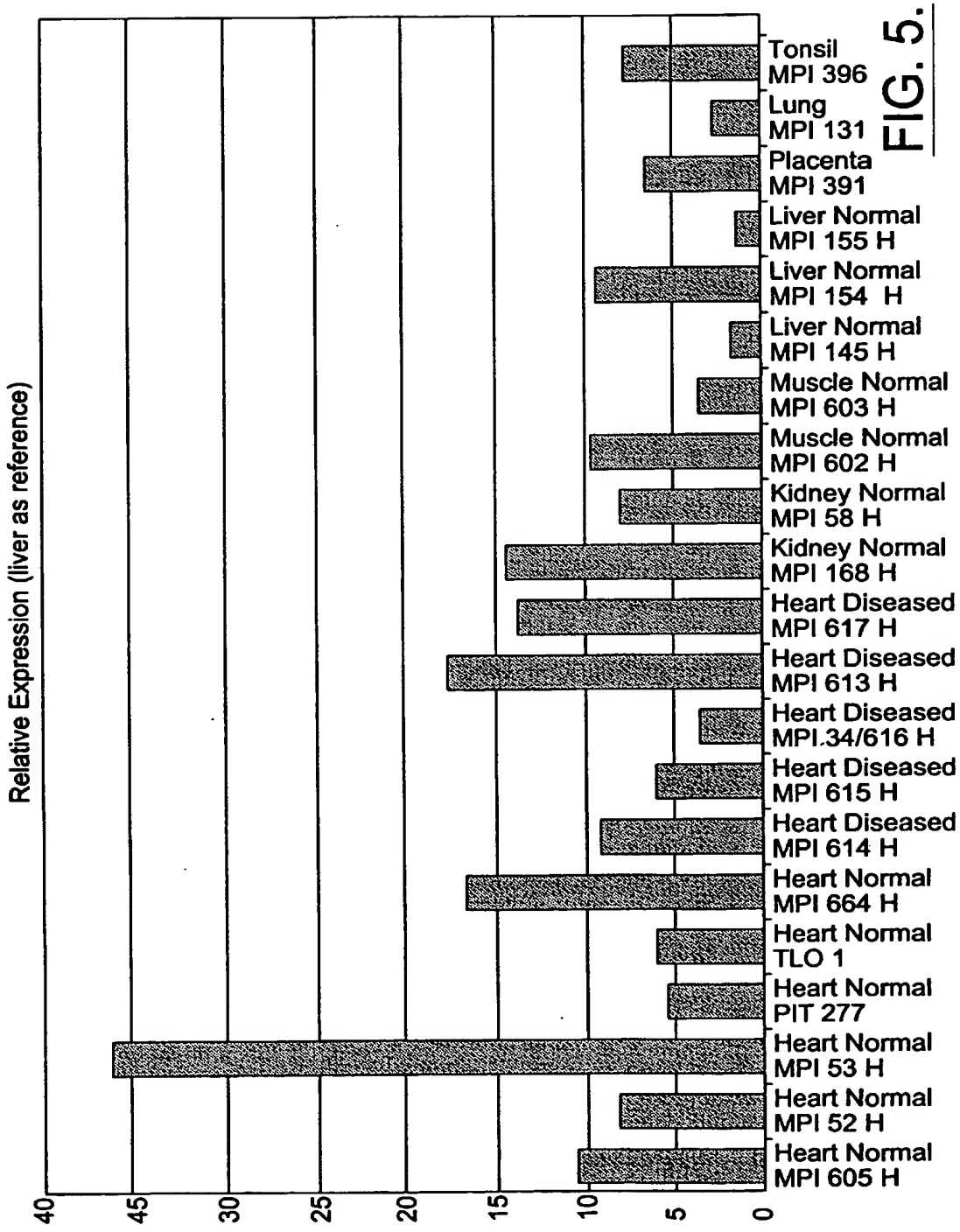

Relative Expression (liver as reference)
40
45
30
25
20
15
10
5
0
Tonsil MPI 396
Lung MPI 131
Placenta MPI 391
Liver Normal MPI 155 H
Liver Normal MPI 154 H
Liver Normal MPI 145 H
Muscle Normal MPI 603 H
Muscle Normal MPI 602 H
Kidney Normal MPI 58 H
Kidney Normal MPI 168 H
Heart Diseased MPI 617 H
Heart Diseased MPI 613 H
Heart Diseased MPI 34/616 H
Heart Diseased MPI 615 H
Heart Diseased MPI 614 H
Heart Normal MPI 664 H
Heart Normal TLO 1
Heart Normal PIT 277
Heart Normal MPI 53 H
Heart Normal MPI 52 H
Heart Normal MPI 605 H

FIG. 5.

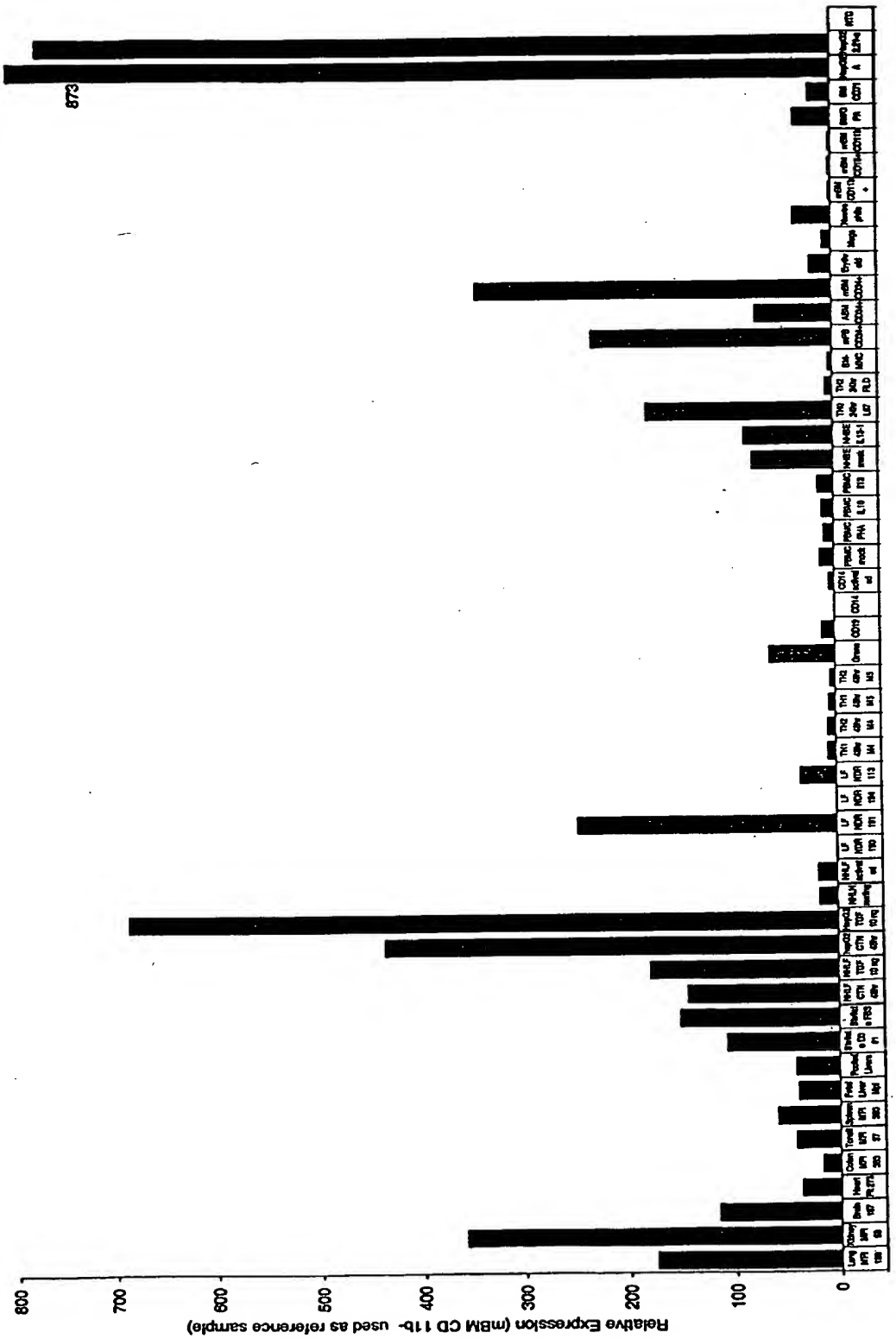
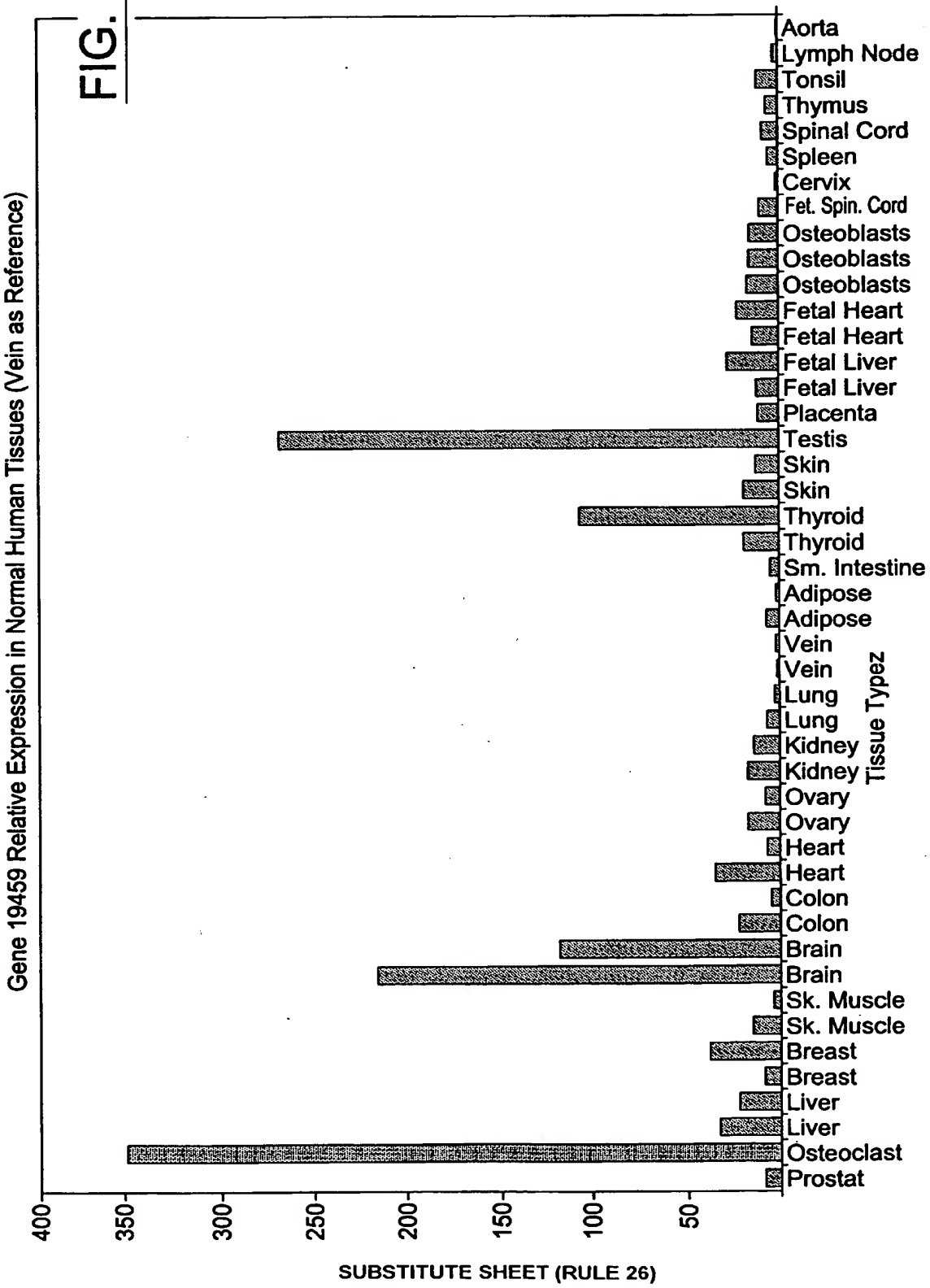

873
800
700
600
500
400
300
200
100
0
Relative Expression (mBM CD 11b- used as reference sample)

FIGURE 6

FIG. 7.


Gene 19459 Relative Expression in Normal Human Tissues (Vein as Reference)
400
350
300
250
200
150
100
50
Tissue Typez
Aorta
Lymph Node
Tonsil
Thymus
Spinal Cord
Spleen
Cervix
Fet. Spin. Cord
Osteoblasts
Osteoblasts
Osteoblasts
Fetal Heart
Fetal Heart
Fetal Liver
Fetal Liver
Placenta
Testis
Skin
Skin
Thyroid
Thyroid
Sm. Intestine
Adipose
Adipose
Vein
Vein
Lung
Lung
Kidney
Kidney
Ovary
Ovary
Heart
Heart
Colon
Colon
Brain
Brain
Sk. Muscle
Sk. Muscle
Breast
Breast
Liver
Liver
Osteoclast
Prostat

SEQUENCE LISTING

<110> Glucksmann, Maria Alexandra
White, David

<120> 19459 Protein, A Novel Seven Transmembrane Protein

<130> 035800/209004

<160> 3

<170> FastSEQ for Windows Version 4.0

<210> 1
<211> 1737
<212> DNA
<213> Homo sapiens

<220>
<221> CDS
<222> (527)...(1717)

<400> 1
ttcggcacga ggcacacctt gggcgtgtgt gggagcagga gagggattct ctggggtccc 60
catttggaac ctggagaccc acttctgcca acaagtcctt cagcctctct gggccttagt 120
ttccccccag tgaagtgggg atagtgacag ctccccgggg agccagtgaa aactttgtga 180
agggagtgat tactgcgggc gaaggcgctt ggcacagcca caactgaagc tcagccgtcg 240
gatgtgatta ttgttgctat tgttaccgct ccgcctggac tgggtgcccg gtcggaaagt 300
cccgctcccc gctcctcctc ccaggcgcgg aagcttcgcg ccgcccgcct cgggggacac 360
ctcccgcgcc atctcggggg cactggagat aaggaaagga acctcctccc tggctgggat 420
ggctgcccgg gcgcgggagg agggccgggg acccggtgag ggaggaccgt agggcccgtg 480
ggggcttcac gcctggcagg cggccggcca ggcctccctc cctgac atg gag agt 535
Met Glu Ser
1

aac ctg tct ggc ctg gtg cct gct gcc ggg ctg gtg cct gcg ctg cca 583
Asn Leu Ser Gly Leu Val Pro Ala Ala Gly Leu Val Pro Ala Leu Pro
5 10 15

cct gct gtg acc ctg ggg ctg aca gct gcc tac acc acc ctg tat gcc 631
Pro Ala Val Thr Leu Gly Leu Thr Ala Ala Tyr Thr Thr Leu Tyr Ala
20 25 30 35

ctg ctc ttc ttc tcc gtc tat gcc cag ctc tgg ctg gtg ctt ctg tat 679
Leu Leu Phe Phe Ser Val Tyr Ala Gln Leu Trp Leu Val Leu Leu Tyr
40 45 50

ggg cac aag cgt ctc agc tat cag acg gtg ttc ctg gcc ctc tgt ctg 727
Gly His Lys Arg Leu Ser Tyr Gln Thr Val Phe Leu Ala Leu Cys Leu
55 60 65

ctc tgg gcc gcc ttg cgt acc acc ctc ttc tcc ttc tac ttc cga gat 775
Leu Trp Ala Ala Leu Arg Thr Thr Leu Phe Ser Phe Tyr Phe Arg Asp
70 75 80

act ccc cgc gcc aac cgc ctg ggg ccc ttg ccc ttc tgg ctt ctc tac 823
Thr Pro Arg Ala Asn Arg Leu Gly Pro Leu Pro Phe Trp Leu Leu Tyr
85 90 95

tgc tgc ccc gtc tgc ctg cag ttc ttc acc ttg acg ctt atg aac ctc 871
Cys Cys Pro Val Cys Leu Gln Phe Phe Thr Leu Thr Leu Met Asn Leu
100 105 110 115

tac ttt gcc cag gtg gtg ttc aag gcc aag gtg aag cgt cgg ccg gag 919
Tyr Phe Ala Gln Val Val Phe Lys Ala Lys Val Lys Arg Arg Pro Glu

```
                120                 125                 130

atg agc cga ggc ttg ctc gct gtc cga ggg gcc ttt gtg ggg gcc tcg      967
Met Ser Arg Gly Leu Leu Ala Val Arg Gly Ala Phe Val Gly Ala Ser
            135                 140                 145

ctg ctc ttt ctg ctg gtg aac gtg ctg tgt gct gtg ctc tcc cat cgg     1015
Leu Leu Phe Leu Leu Val Asn Val Leu Cys Ala Val Leu Ser His Arg
        150                 155                 160

cgc cgg gca cag ccc tgg gcc ctg ctg ctt gtc cgc gtc ctg gtg agc     1063
Arg Arg Ala Gln Pro Trp Ala Leu Leu Leu Val Arg Val Leu Val Ser
    165                 170                 175

gac tcc ctg ttc gtc atc tgc gcg ctg tct ctt gct gcc tgc ctc tgc     1111
Asp Ser Leu Phe Val Ile Cys Ala Leu Ser Leu Ala Ala Cys Leu Cys
180                 185                 190                 195

ctc gtc gcc agg cgg gcg ccc tcc act agc atc tac ctg gag gcc aag     1159
Leu Val Ala Arg Arg Ala Pro Ser Thr Ser Ile Tyr Leu Glu Ala Lys
                200                 205                 210

ggg acc agt gtg tgc cag gcg gcc gcg atg ggt ggc gcc atg gtc ctg     1207
Gly Thr Ser Val Cys Gln Ala Ala Ala Met Gly Gly Ala Met Val Leu
            215                 220                 225

ctc tat gcc agc cgg gcc tgc tac aac ctg aca gca ctg gcc ttg gcc     1255
Leu Tyr Ala Ser Arg Ala Cys Tyr Asn Leu Thr Ala Leu Ala Leu Ala
        230                 235                 240

ccc cag agc cgg ctg gac acc ttc gat tac gac tgg tac aat gtg tct     1303
Pro Gln Ser Arg Leu Asp Thr Phe Asp Tyr Asp Trp Tyr Asn Val Ser
    245                 250                 255

gac cag gcg gac ctg gtg aat gac ctg ggg aac aaa ggc tac ctg gta     1351
Asp Gln Ala Asp Leu Val Asn Asp Leu Gly Asn Lys Gly Tyr Leu Val
260                 265                 270                 275

ttt ggc ctc atc ctc ttc gtg tgg gag cta ctg ccc acc acc ctg ctg     1399
Phe Gly Leu Ile Leu Phe Val Trp Glu Leu Leu Pro Thr Thr Leu Leu
                280                 285                 290

gtg ggc ttc ttc cgg gtg cac cgg ccc cca cag gac ctg agc acc agc     1447
Val Gly Phe Phe Arg Val His Arg Pro Pro Gln Asp Leu Ser Thr Ser
            295                 300                 305

cac atc ctc aat ggg cag gtc ttt gcc tct cgg tcc tac ttc ttt gac     1495
His Ile Leu Asn Gly Gln Val Phe Ala Ser Arg Ser Tyr Phe Phe Asp
        310                 315                 320

cgg gct ggg cac tgt gaa gat gag ggc tgc tcc tgg gag cac agc cgg     1543
Arg Ala Gly His Cys Glu Asp Glu Gly Cys Ser Trp Glu His Ser Arg
    325                 330                 335

ggt gag agc acc agt atg tcg ggc agt cta ggc tct ggg agc tgg tat     1591
Gly Glu Ser Thr Ser Met Ser Gly Ser Leu Gly Ser Gly Ser Trp Tyr
340                 345                 350                 355

ggt gcc atc ggg cgt gag ccg ggc tgg tat ggg ggc agc cag acg aag     1639
Gly Ala Ile Gly Arg Glu Pro Gly Trp Tyr Gly Gly Ser Gln Thr Lys
                360                 365                 370

acc act cct ctg ctc ttc tcc cag gtg cca gga cca ggc ggc cac cac     1687
Thr Thr Pro Leu Leu Phe Ser Gln Val Pro Gly Pro Gly Gly His His
            375                 380                 385

cac agt ctc tac tcc acc cca cag acg tga tccccccctgg tgccgaattc      1737
```

```
His Ser Leu Tyr Ser Thr Pro Gln Thr  *
        390                 395


<210> 2
<211> 396
<212> PRT
<213> Homo sapiens

<400> 2
Met Glu Ser Asn Leu Ser Gly Leu Val Pro Ala Ala Gly Leu Val Pro
 1               5                  10                  15
Ala Leu Pro Pro Ala Val Thr Leu Gly Leu Thr Ala Ala Tyr Thr Thr
            20                  25                  30
Leu Tyr Ala Leu Leu Phe Phe Ser Val Tyr Ala Gln Leu Trp Leu Val
        35                  40                  45
Leu Leu Tyr Gly His Lys Arg Leu Ser Tyr Gln Thr Val Phe Leu Ala
    50                  55                  60
Leu Cys Leu Leu Trp Ala Ala Leu Arg Thr Thr Leu Phe Ser Phe Tyr
65                  70                  75                  80
Phe Arg Asp Thr Pro Arg Ala Asn Arg Leu Gly Pro Leu Pro Phe Trp
                85                  90                  95
Leu Leu Tyr Cys Cys Pro Val Cys Leu Gln Phe Phe Thr Leu Thr Leu
            100                 105                 110
Met Asn Leu Tyr Phe Ala Gln Val Val Phe Lys Ala Lys Val Lys Arg
        115                 120                 125
Arg Pro Glu Met Ser Arg Gly Leu Leu Ala Val Arg Gly Ala Phe Val
    130                 135                 140
Gly Ala Ser Leu Leu Phe Leu Leu Val Asn Val Leu Cys Ala Val Leu
145                 150                 155                 160
Ser His Arg Arg Arg Ala Gln Pro Trp Ala Leu Leu Leu Val Arg Val
                165                 170                 175
Leu Val Ser Asp Ser Leu Phe Val Ile Cys Ala Leu Ser Leu Ala Ala
            180                 185                 190
Cys Leu Cys Leu Val Ala Arg Arg Ala Pro Ser Thr Ser Ile Tyr Leu
        195                 200                 205
Glu Ala Lys Gly Thr Ser Val Cys Gln Ala Ala Ala Met Gly Gly Ala
    210                 215                 220
Met Val Leu Leu Tyr Ala Ser Arg Ala Cys Tyr Asn Leu Thr Ala Leu
225                 230                 235                 240
Ala Leu Ala Pro Gln Ser Arg Leu Asp Thr Phe Asp Tyr Asp Trp Tyr
                245                 250                 255
Asn Val Ser Asp Gln Ala Asp Leu Val Asn Asp Leu Gly Asn Lys Gly
            260                 265                 270
Tyr Leu Val Phe Gly Leu Ile Leu Phe Val Trp Glu Leu Leu Pro Thr
        275                 280                 285
Thr Leu Leu Val Gly Phe Phe Arg Val His Arg Pro Pro Gln Asp Leu
    290                 295                 300
Ser Thr Ser His Ile Leu Asn Gly Gln Val Phe Ala Ser Arg Ser Tyr
305                 310                 315                 320
Phe Phe Asp Arg Ala Gly His Cys Glu Asp Glu Gly Cys Ser Trp Glu
                325                 330                 335
His Ser Arg Gly Glu Ser Thr Ser Met Ser Gly Ser Leu Gly Ser Gly
            340                 345                 350
Ser Trp Tyr Gly Ala Ile Gly Arg Glu Pro Gly Trp Tyr Gly Gly Ser
        355                 360                 365
Gln Thr Lys Thr Thr Pro Leu Leu Phe Ser Gln Val Pro Gly Pro Gly
    370                 375                 380
Gly His His His Ser Leu Tyr Ser Thr Pro Gln Thr
385                 390                 395

<210> 3
<211> 1191
<212> DNA
<213> Homo sapiens

<400> 3
```

```
atggagagta acctgtctgg cctggtgcct gctgccgggc tggtgcctgc gctgccacct      60
gctgtgaccc tggggctgac agctgcctac accaccctgt atgccctgct cttcttctcc     120
gtctatgccc agctctggct ggtgcttctg tatgggcaca agcgtctcag ctatcagacg     180
gtgttcctgg ccctctgtct gctctgggcc gccttgcgta ccaccctctt ctccttctac     240
ttccgagata ctccccgcgc caaccgcctg gggcccttgc ccttctggct tctctactgc     300
tgccccgtct gcctgcagtt cttcaccttg acgcttatga acctctactt tgcccaggtg     360
gtgttcaagg ccaaggtgaa gcgtcggccg gagatgagcc gaggcttgct cgctgtccga     420
ggggcctttg tgggggcctc gctgctcttt ctgctggtga acgtgctgtg tgctgtgctc     480
tcccatcggc gccgggcaca gccctgggcc ctgctgcttg tccgcgtcct ggtgagcgac     540
tccctgttcg tcatctgcgc gctgtctctt gctgcctgcc tctgcctcgt cgccaggcgg     600
gcgccctcca ctagcatcta cctggaggcc aaggggacca gtgtgtgcca ggcggccgcg     660
atgggtggcg ccatggtcct gctctatgcc agccgggcct gctacaacct gacagcactg     720
gccttggccc cccagagccg gctggacacc ttcgattacg actggtacaa tgtgtctgac     780
caggcggacc tggtgaatga cctggggaac aaaggctacc tggtatttgg cctcatcctc     840
ttcgtgtggg agctactgcc caccaccctg ctggtgggct tcttccgggt gcaccggccc     900
ccacaggacc tgagcaccag ccacatcctc aatgggcagg tctttgcctc tcggtcctac     960
ttctttgacc gggctgggca ctgtgaagat gagggctgct cctgggagca cagccggggt    1020
gagagcacca gtatgtcggg cagtctaggc tctgggagct ggtatggtgc catcgggcgt    1080
gagccgggct ggtatggggg cagccagacg aagaccactc ctctgctctt ctcccaggtg    1140
ccaggaccag gcggccacca ccacagtctc tactccaccc cacagacgtg a              1191
```